

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:32:14 ; Search time 34.23 Seconds
(without alignments)
342.345 Million cell updates/sec

Title: US-09-466-935-2
Perfect score: 1061
Sequence: 1 MLEWWEFAYLLTSLITLSP.....KIFGSLPMVGLIASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues
Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0401:*
1: /SID52/gcgdata/geneseq/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1061	100.0	205	Y99597	E. coli L-homoseri
2	1061	100.0	205	Y79298	E. coli RhtB prote
3	195.5	18.4	212	B01787	Escherichia coli y
4	155	14.6	153	R87527	Mel-linked mlgA ge
5	155	14.6	153	W73358	S. colwelliana Mlg
6	144	13.6	130	B40407	Human OREF ORF171
7	126.5	11.9	223	B01786	Escherichia coli y
8	126	11.9	195	B01788	Escherichia coli y
9	123.5	11.6	229	B11637	A. yllis hypersens
10	122	11.5	206	Y99598	E. coli L-threonin
11	113	10.7	211	B01789	Escherichia coli y

12	89.5	8.4	236	18	W37715	C. glutamicum Lys
13	80	7.5	484	20	Y04104	Arthrobacter nicot
14	80	7.5	517	20	Y04105	Arthrobacter nicot
15	79.5	7.5	307	20	Y38701	Neisseria gonorrhoe
16	79.5	7.5	307	20	Y38702	Neisseria gonorrhoe
17	79	7.4	596	20	Y09518	C. elegans dopamin
18	79	7.4	615	20	Y09517	C. elegans dopamin
19	78.5	7.4	513	20	Y38788	Neisseria gonorrhoe
20	78.5	7.4	513	20	Y38789	Neisseria gonorrhoe
21	77	7.3	496	18	W20694	H. pylori putative
22	77	7.3	539	18	W20694	H. pylori putative
23	76.5	7.2	213	21	G51077	Arabidopsis thalia
24	76.5	7.2	242	21	G51076	Arabidopsis thalia
25	76.5	7.2	243	21	G12759	Arabidopsis thalia
26	76.5	7.2	246	21	G51075	Arabidopsis thalia
27	76.5	7.2	272	21	G12758	Arabidopsis thalia
28	76.5	7.2	276	21	G12757	Arabidopsis thalia
29	76	7.2	553	20	Y01651	Arabidopsis thalia
30	76	7.2	553	21	B20579	A protein with cat
31	75.5	7.1	277	20	Y38699	Mouse OCTN1 amino
32	75.5	7.1	396	17	W02112	Neisseria meningit
33	75.5	7.1	396	17	W02113	Human homologue of
34	74.5	7.0	170	20	Y29197	Murine Int6 protei
35	74.5	7.0	465	20	Y35106	Amino acid sequenc
36	74	7.0	277	20	Y38698	Chlamydia pneumoni
37	74	7.0	551	20	Y01649	Neisseria meningit
38	74	7.0	783	21	Y96164	A protein with cat
39	73.5	6.9	153	19	W80604	S. typhimurium pox
40	73.5	6.9	375	18	W20731	H. pylori inner me
41	73.5	6.9	955	21	B16338	Eucalyptus grandis
42	73.5	6.9	974	19	W73308	Cellulose synthase
43	73.5	6.9	974	19	W60039	Cotton cellulose s
44	73	6.9	362	21	B01449	Rodent GPR2 protei
45	72	6.8	468	20	Y34877	Amino acid sequenc

ALIGNMENTS

RESULT 1	
ID	Y99597 standard; Protein: 205 AA.
Y99597	
XX	
AC	Y99597:
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	E. coli L-homoserine resistance protein, RhtB.
XX	
KW	L-homoserine resistance; L-homoserine synthesis; rhtB;
KW	L-threonine; L-valine; L-leucine.
XX	
OS	Escherichia coli.
XX	
PN	EP1013765-A1.
XX	
PD	28-JUN-2000.
XX	
PF	20-DEC-1999; 99EP-0125406.
XX	
PR	23-DEC-1998; 98RU-0123511.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
PI	Livshits VA, Zakaraeva NP, Aleshin VV, Belareva AV, Tokmakova IL;
XX	
DR	WPI: 2000-414602/36.
XX	
DR	N-PSDB; A48442.
XX	
PT	Novel Escherichia bacterium having enhanced L-threonine resistance due
PT	to enhanced RhtC protein activity, used to produce L-threonine,
PT	L-homoserine, L-valine and L-leucine -
XX	

PS Claim 2; Page 13-14; 2App; English.
 XX
 CC The present sequence is the L-homoserine resistance protein, RhtB, from
 CC Escherichia coli. The coding sequence may be used to impart L-homoserine
 CC resistance on E. coli bacteria, which would be useful for producing
 CC a high yield of L-homoserine. L-homoserine resistance means that the
 CC bacteria will be able to grow on a minimal medium containing
 CC L-homoserine at a concentration at which the corresponding wild-type
 CC strain would not grow. Since the transformed bacteria can grow on the
 CC minimal medium, it can synthesise L-homoserine, which accumulates. The
 CC accumulated amino acids can then be removed from the culture medium.
 CC The bacterium of the present invention may also be used to synthesise
 CC L-threonine, L-valine and L-leucine at increased levels.
 XX
 SQ Sequence 205 AA;
 Query Match 100.0%; Score 1061; DB 21; Length 205;
 Best Local Similarity 100.0%; Pred. No. 3.9e-113;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLEWFWFAYLTLITLSPSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGRGV 60
 DB 1 mtlewfwaylltlltllspsgaintmttshygpaggyvwadrtgdsycagrgv 60
 QY 61 GILFSSVIAFEYKWKAGAAYLIMIGIQWRRAAGATDKSLASTOSRRHLFORAVVNL 120
 DB 61 gtlfssvიაfevkwagaaayliwlgigwraagaidklsastgrrhlfgravvnl 120
 QY 121 NPKSIVFLAALFPQFTMPQOPOLMOYIVLGVTIIVDIIIMIGYATLAORIALMIGPKQ 180
 DB 121 npksivflaalfpqftimpqopqimgyivlgvtliivdliimigyalraqrialwtkpqr 180
 QY 181 MKALNKIFGSLFMLVGALLASARHA 205
 DB 181 mkalnki fgs lfmlygallasarha 205
 RESULT 2
 Y79298
 ID Y79298 standard; Protein; 205 AA.
 AC Y79298;
 XX
 DT 18-JUL-2000 (first entry)
 DE E. coli RhtB protein imparting homoserine resistance.
 XX
 KW Homoserine resistance; RhtB protein; L-homoserine; L-alanine;
 KW L-isoleucine; L-valine; L-threonine.
 XX
 OS Escherichia coli.
 XX
 PN EP994190-A2.
 PD 19-APR-2000.
 XX
 PF 20-SEP-1999; 99EP-0118581.
 XX
 PR 13-OCT-1998; 98RU-0118425.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI livshits VA, zakataeva NP, Aleoshin VV, Belareova AV;
 PI Tokmakova IL;
 DR WPI: 2000-273530/24.
 DR N-PSDB: Z94405.
 XX
 PT Novel RhtB protein, useful for generation of L-homoserine resistance in
 PT Escherichia bacteria and large-scale production of e.g. L-homoserine
 PT and L-alanine -
 XX

PS Claim 1; Page 11-12; 14pp; English.
 XX
 CC The present sequence is that of the novel Escherichia coli K-12
 CC protein, RhtB, which participates in resistance to L-homoserine.
 CC Amplification of the rhtB gene (see Z94405) results in an
 CC improvement of the amino acid productivity of E. coli. The
 CC invention provides: a protein comprising the present sequence or 1
 CC including a deletion, substitution, insertion and/or addition of 1
 CC or more amino acids and having the activity of making a bacterium
 CC resistant to L-homoserine; DNA encoding RhtB; a bacterium,
 CC especially of the genus Escherichia, in which L-homoserine
 CC resistance is enhanced by amplifying the copy number or increasing
 CC the expression rate of the rhtB DNA, the DNA being carried on a
 CC multicopy vector or on a transposon; and a method for producing an
 CC amino acid by cultivating the bacterium in a culture medium to
 CC produce and accumulate the amino acid in the medium, from which it
 CC is recovered. The method is used for the production of
 CC L-homoserine, L-alanine, L-isoleucine, L-valine or L-threonine (all
 CC claimed).
 XX
 SQ Sequence 205 AA;
 Query Match 100.0%; Score 1061; DB 21; Length 205;
 Best Local Similarity 100.0%; Pred. No. 3.9e-113;
 Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLEWFWFAYLTLITLSPSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGRGV 60
 DB 1 mtlewfwaylltlltllspsgaintmttshygpaggyvwadrtgdsycagrgv 60
 QY 61 GILFSSVIAFEYKWKAGAAYLIMIGIQWRRAAGATDKSLASTOSRRHLFORAVVNL 120
 DB 61 gtlfssvიაfevkwagaaayliwlgigwraagaidklsastgrrhlfgravvnl 120
 QY 121 NPKSIVFLAALFPQFTMPQOPOLMOYIVLGVTIIVDIIIMIGYATLAORIALMIGPKQ 180
 DB 121 npksivflaalfpqftimpqopqimgyivlgvtliivdliimigyalraqrialwtkpqr 180
 QY 181 MKALNKIFGSLFMLVGALLASARHA 205
 DB 181 mkalnki fgs lfmlygallasarha 205
 RESULT 3
 B01787
 ID B01787 standard; Protein; 212 AA.
 AC B01787;
 XX
 DT 03-JAN-2001 (first entry)
 DE Escherichia coli Yeas amino acid excretion protein.
 XX
 KW E. coli; yeas gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein.
 XX
 OS Escherichia coli.
 XX
 PN EP1016710-A2.
 PD 05-JUL-2000.
 XX
 PF 17-DEC-1999; 99EP-0125263.
 XX
 PR 30-DEC-1998; 98RU-0124016.
 PR 09-MAR-1999; 99RU-0104431.
 XX
 PA (AJIN) AJINOMOTO CO INC.
 XX
 PI livshits VA, zakataeva NP, Nakanishi K, Aleoshin VV, Troshin PV;
 PI Tokmakova IL;
 XX

DR	WP1: 2000-414802/36.
DR	N-PSDB; A52689.
XX	
PT	Increased production of L-amino acids by an Escherichia bacterium
XX	comprises increasing the expression amount of an L-amino acid excretion
XX	protein -
XX	
PS	Claim 1; Page 21; 29pp; English.
CC	
CC	The present sequence is the Yeast amino acid excretion protein from
CC	Escherichia coli. This protein is involved in the production of amino
CC	acids, and an increase in its expression leads to an increased
CC	accumulation of amino acids in the cell. In this case, an increase in
CC	lysine, alanine, valine, histidine, isoleucine, glutamic acid and proline
CC	is achieved if multiple copies of its gene are transfected into a
CC	bacterium. The bacterium used is E. coli.
XX	
SQ	Sequence 212 AA;
XX	
Query Match	18.4%; Score 195.5; DB 21; Length 212;
Best Local Similarity	28.6%; Pred. No. 2.4e-14;
Matches 61; Conservative 45; Mismatches 82; Indels 25; Gaps 10.	
OY	3 LEWMAVLLTSLILTPSSGAI---NTMTSLNHGYAG-GVYCWSADRTGSYC--A 55 : : : : : : : : : : : Db lnyv-tylvgaifivlyvpgnlffvlknsvsqmkgyilaacyf-----lgdavlmlf 60
OY	56 GWRCVGTLFRSRYIAEVLKMGAAATLMLGIQQWRAGAIDLR---SLASTOSRRH--L 110 : : : : : : : : : Db 61 awagvatllktpllnlvrylgafilylgskllyal---lkgnkseaksdepygal 116
OY	111 FQRVFNLTNPKSIVFELALFPQFIIMPQPOL-MQYIVLGVTITVDIYMIGYATIAD 169 : : : : : : : : : : : : : : : Db 117 fkrallstlnpallfyvsffvgfdvnaphgtglsffllaetlelvsfcylsfllisga 176
OY	170 RIALWKGRQM-KALKTIFGSEFMVGCALLAS 201 : : : : : : : : : : : : : : : Db 177 ftvtgirtkkklakvgsnlglmfvgaarlat 209
RESULT 4	
R87527	
ID	R87527 standard; Protein: 153 AA.
XX	
AC	R87527;
XX	
DT	01-JUL-1996 (first entry)
XX	
DE	Mel-linked mlga gene product.
XX	
KW	Marine melA; selectable marker; oyster larva settlement;
KW	pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
KW	mlga gene.
XX	
OS	Shewanella colwelliana strain LSTDYF.
XX	
PN	US5474933-A.
XX	
PD	12-DEC-1995.
XX	
PX	21-MAR-1990; 90US-0496804.
XX	
PR	08-NOV-1993; 93US-0148945.
XX	
PR	21-MAR-1990; 90US-0496804.
XX	
PR	10-NOV-1992; 92US-0974837.
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fugua WC, Weiner RM;
XX	
XX	WP1: 1996-039515/04.
DR	N-PSDB; T06767.

[illegible]


```

XX 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
XX (AJIN ) AJINOMOTO CO INC.
PA
PI livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
DR WPI: 2000-414802/36.
DR N-PSDB; A52691.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
XX comprises increasing the expression amount of an L-amino acid excretion
XX protein -
XX
XX Claim 1; Page 25; 29pp; English.
XX
XX The present sequence is the yggA amino acid excretion protein from
XX Escherichia coli. This protein is involved in the production of amino
XX acids, and an increase in its expression leads to an increased
XX accumulation of amino acids in the cell. In this case, an increase in
XX arginine, glutamic acid and lysine is achieved if multiple copies of the
XX gene are transfected into a bacterium. The bacterium used is E. coli.
XX
SQ Sequence 211 AA:

Query Match 10.7%; Score 113; DB 21; Length 211;
Best Local Similarity 24.9%; Pred. No. 6.2e-05;
Matches 52; Conservative 36; Mismatches 95; Indels 26; Gaps 10;

QY 7 FAY-----LITSLITLSPGSGAINTMTSLNHYGYPAG-GVYCWASDRTGDSYCGMRG 59
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 tsyyfgqlaigaa mlipqpna fy--mmgqitrrqymaia lca lsdv--l lca g lfy 57

QY 60 VGFPSRSVIAFEYLMKAGAAVLIMLGIOQWRAGAIDLKSLAST-----QSSRHLPORAV 115
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 58 gsal lmgspwllalvclwgvafliwyfgafk tamsa nie-lasevmkqgwt klalm l 116

QY 116 EVNLTNP-----KSTFLALAFPOFIMPOQPOLMOXYIVLGVTITVDIIMIGATLAQRI 171
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 avtclnphvy l dcfv l v l g s l 99g--l dvepr--t w falg--l t s a s f l w f g l a l a a w l 170

QY 172 ALMTKGPCKAKLNTKIFGSLFEMVGAALIA 200
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 171 aprlrltakgrlln l v g c m w f l a l g l a 199

RESULT 12
W37715
ID W37715 standard; Protein; 236 AA.
XX
XX W37715.
XX
XX 12-MAR-1998 (first entry)
XX
XX C. glutamicum Lys E protein (lysine export protein).
XX
XX LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
XX Microbial production; amino acid; animal feed additive.
XX
XX Corynebacterium glutamicum.
XX
XX DE19548222-A1.
XX
XX 26-JUN-1997.
XX
XX 22-DEC-1995; 95DE-1048222.
XX
XX 22-DEC-1995; 95DE-1048222.
XX
XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.
PA

```

```

XX Eggeling L, Sahm H, Vrljic M;
XX
XX WPI: 1997-333867/31.
XX
XX N-PSDB; T96816.
XX
XX Increasing microbial production of amino acids, especially lysine -
XX by improving export carrier activity or corresponding gene
XX expression, also new export and regulatory genes from
XX Corynebacterium
XX
XX Claim 42; Page 10; 16pp; German.
XX
XX This sequence is the LysE protein product, a lysine export protein.
XX LysG and LysE encode a lysine transport regulatory protein and an export
XX protein, respectively. Microbial production of amino acids (A) is
XX improved by increasing the export-carrier activity and/or the export gene
XX expression in a microorganism that produces (A). The method is
XX specifically used to increase production of lysine, used as an animal
XX feed additive. Other (A) are variously useful as pharmaceuticals,
XX condiments and intermediates for fine chemicals. This method increases
XX the amount of (A) secreted into the culture medium. Export of (A) has
XX been found to depend on a single gene.
XX
SQ Sequence 236 AA:

Query Match 8.4%; Score 89.5; DB 16; Length 236;
Best Local Similarity 20.6%; Pred. No. 0.035;
Matches 36; Conservative 25; Mismatches 79; Indels 35; Gaps 4;

QY 3 LEWMPAYLITSLITLSPGSGAINTMTSL-NHGYPAAGVYCWASDRTGDSYCGMRG 61
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 4 melficglllgaslllsigpqnvlv i k g l k r e g l i a v l l v c l s d v f--l f l e g l v g d 61

QY 62 TLFPSRSVIAFEYLMKAGAAVLIMLGIOQWRAGAIDLKSLASTQSRHLPORAVFVNLTN 121
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 l l s n a a p l v l d m r v g y l a j l l w f a m a a k d a -----m t n 96

QY 122 KSTFLALAFPOFIMPOQPOLMOXYIVLGVTITVDIIMIGATLAQRIALMTK 176
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 kvea-----p q l l e e c e p v p d d t p l g s a v a t d t m r v e v s v d k g r v w k 144

RESULT 13
Y04104
ID Y04104 standard; Protein; 484 AA.
XX
XX Y04104;
XX
XX 10-JUN-1999 (first entry)
XX
XX Arthrobacter nicotinovorans levan fructotransferase protein #1.
XX
XX Arthrobacter nicotinovorans; levan fructotransferase.
XX
XX Arthrobacter nicotinovorans.
XX
XX JP11069978-A.
XX
XX 16-MAR-1999.
XX
XX 28-AUG-1997; 97JP-0232421.
XX
XX 28-AUG-1997; 97JP-0232421.
XX
XX WPI: 1999-247463/21.
XX
XX N-PSDB; X19826.
XX
XX Levan fructotransferase gene - for recombinant production of levan
XX fructotransferase
PA

```

XX Claim 1; Page 5-8; 14pp; Japanese.

XX The present sequence represents *Arthrobacter nicotinovorus* levan
CC fructotransferase. The present invention also describes a method
CC for the preparation of levan fructotransferase in which a transformant
CC is cultured in a medium and levan fructotransferase is collected from
CC the culture. The method can prepare levan fructotransferase in a
CC large amount.

XX Sequence 484 AA;

Query Match 7.5%; Score 80; DB 20; Length 484;
Best Local Similarity 19.3%; Pred. No. 1.1;
Matches 41; Conservative 34; Mismatches 81; Indels 56; Gaps 7;

OY 3 LEW---WFA-----YLLTSIIITLSPSGAINTMTSLNHG 35
DB 254 ldwgdwyaaavtwpaveapetkrlataamnnkyaarnvptdasdyngqnsitrelile 313
OY 36 YPAGGVYC-----WASDRITGDSYCAGMRGVTLFSRSYIAFEVILKAGAAVLIW 84
DB 314 rqsqgwyllstlpypalsnyatset-----llpdrtrngsfvlpwsgrayele 361
OY 85 LGIQOMRAGAIDDKSLASTOSRRHL---FQRAVFVNLTPKSIYVFLAALFPQETMPQO 140
DB 362 ldl-swtdaanvsgvrsdgsrhtnlgkygdelydrasseqsgyalapycraapid 420
OY 141 POLMQ-YIVLGVTITVVDIIVMIGVATLAORI 171
DB 421 anarsvhlrlftvdtsqevfvsnghtvsgqv 452

RESULT 14
Y04105
ID Y04105 standard; Protein: 517 AA.

XX Y04105;

DT 10-JUN-1999 (first entry)

XX *Arthrobacter nicotinovorus* levan fructotransferase protein #2.

DE *Arthrobacter nicotinovorus*; levan fructotransferase.

OS *Arthrobacter nicotinovorus*.

XX JP1069978-A.

PN 16-MAR-1999.

PD 28-AUG-1997; 97JP-0232421.

PR 28-AUG-1997; 97JP-0232421.

PA (NIOC) NIPPON OIL CO LTD.

DR WPI: 1999-247463/21.

XX N-PSDB; X19827.

PT Levan fructotransferase gene - for recombinant production of levan

XX fructotransferase

PS Claim 3; Page 8-9; 14pp; Japanese.

XX The present sequence represents *Arthrobacter nicotinovorus* levan
CC fructotransferase. The present invention also describes a method
CC for the preparation of levan fructotransferase in which a transformant
CC is cultured in a medium and levan fructotransferase is collected from
CC the culture. The method can prepare levan fructotransferase in a
CC large amount.

SQ Sequence 517 AA;

Query Match 7.5%; Score 80; DB 20; Length 517;
Best Local Similarity 19.3%; Pred. No. 1.2;
Matches 41; Conservative 34; Mismatches 81; Indels 56; Gaps 7;

OY 3 LEW---WFA-----YLLTSIIITLSPSGAINTMTSLNHG 35
DB 287 ldwgdwyaaavtwpaveapetkrlataamnnkyaarnvptdasdyngqnsitrelile 346
OY 36 YPAGGVYC-----WASDRITGDSYCAGMRGVTLFSRSYIAFEVILKAGAAVLIW 84
DB 347 rqsqgwyllstlpypalsnyatset-----llpdrtrngsfvlpwsgrayele 394
OY 85 LGIQOMRAGAIDDKSLASTOSRRHL---FQRAVFVNLTPKSIYVFLAALFPQETMPQO 140
DB 395 ldl-swtdaanvsgvrsdgsrhtnlgkygdelydrasseqsgyalapycraapid 453
OY 141 POLMQ-YIVLGVTITVVDIIVMIGVATLAORI 171
DB 454 anarsvhlrlftvdtsqevfvsnghtvsgqv 495

RESULT 15
Y38701
ID Y38701 standard; Protein: 307 AA.

XX Y38701;

DT 08-OCT-1999 (first entry)

DE *Neisseria gonorrhoeae* antigen encoded by partial ORF104.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;

XX treatment; *Neisseria* infection; meningitis; septicemia; gonorrhea.

OS *Neisseria gonorrhoeae*.

PN W09924578-A2.

PD 20-MAY-1999.

PE 09-OCT-1998; 98WC-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

PA (CHIR-) CHIRON SPA.

PI Grandi G, Maignani V, Pizze M, Rappuoli R, Scarlato V;

XX WPI: 1999-327407/27.

PT proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for

XX diagnosis, treatment and prevention of infection

PS Claim 4; Page 252; 524pp; English.

XX Amino acid sequences Y38499-Y38944 represent *Neisseria meningitidis*
CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open
CC reading frames (ORFs) 211972-212358. The antigenic proteins
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
CC infections, such as meningitis, septicemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:32:29 ; Search time 21.59 Seconds
(without alignments)
182,410 Million cell updates/sec

Title: US-09-466-935-2
Perfect score: 1061
Sequence: 1 MTELEMPAVLITSLITLSP.....KIFGSLFMLVGLASARHA 205

Scoring table: BLOSUM62
GapOP 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents,AA:*
1: /cgcn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgcn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgcn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgcn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgcn2_6/ptodata/2/1aa/PCYUS.COMB.pep.*
6: /cgcn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155	14.6	153	2	US-08-476-254-7
2	155	14.6	153	6	5474933-4
3	72	6.8	503	4	US-09-068-195-24
4	70	6.6	428	6	5432081-9
5	68.5	6.5	534	4	US-08-875-223-8
6	68.5	6.5	857	2	US-08-448-603A-30
7	68.5	6.5	857	3	US-09-134-075-30
8	68	6.4	456	6	5432081-7
9	67.5	6.4	1302	1	US-08-232-537-2
10	66.5	6.3	291	1	US-08-358-171-2
11	66.5	6.3	291	3	US-09-090-947-2
12	66.5	6.3	344	2	US-08-468-812-2
13	66.5	6.3	898	3	US-08-863-102-1
14	66.5	6.3	898	3	US-08-863-102-4
15	66	6.2	772	1	US-08-802-141-4
16	66	6.2	1098	3	US-08-726-214-10
17	65.5	6.2	379	1	US-08-227-108-18
18	65.5	6.2	379	2	US-09-073-674-18
19	64.5	6.1	380	2	US-08-472-659-34
20	64.5	6.1	380	2	US-08-474-661-34
21	64.5	6.1	380	2	US-08-611-977-34
22	64.5	6.1	453	2	US-08-244-205-13
23	64.5	6.1	453	5	PCT-US92-10284-13
24	64.5	6.1	489	2	US-08-752-307B-3
25	64.5	6.1	530	2	US-08-752-307B-2
26	64.5	6.1	642	1	US-08-706-936-3
27	64	6.0	1052	2	US-08-852-806-2

28	64	6.0	1052	4	US-09-163-669-2	Sequence 2, Appl1
29	63.5	6.0	426	6	5268463-8	Patent No. 5268463
30	63.5	6.0	506	3	US-08-867-352-21	Sequence 21, Appl1
31	63	5.9	371	4	US-09-043-937A-4	Sequence 4, Appl1
32	63	5.9	533	1	US-08-368-071-10	Sequence 10, Appl1
33	63	5.9	533	1	US-08-458-181-10	Sequence 10, Appl1
34	63	5.9	533	5	PCN-US93-02172-10	Sequence 10, Appl1
35	63	5.9	607	1	US-07-959-943-7	Sequence 7, Appl1
36	63	5.9	607	1	US-07-879-617A-12	Sequence 12, Appl1
37	63	5.9	607	1	US-08-753-985-12	Sequence 12, Appl1
38	63	5.9	630	1	US-07-959-943-9	Sequence 9, Appl1
39	63	5.9	642	1	US-08-706-936-2	Sequence 2, Appl1
40	63	5.9	653	1	US-07-782-298-2	Sequence 2, Appl1
41	63	5.9	686	3	US-09-306-922-2	Sequence 2, Appl1
42	63	5.9	759	2	US-08-637-759B-89	Sequence 89, Appl1
43	63	5.9	759	3	US-08-871-355A-89	Sequence 89, Appl1
44	62.5	5.9	197	4	US-09-402-668-2	Sequence 2, Appl1
45	62.5	5.9	280	2	US-08-700-013B-15	Sequence 15, Appl1

ALIGNMENTS

```
RESULT 1
US-08-476-254-7
; Sequence 7, Application US/08476254
; Patent No. 5846531
; GENERAL INFORMATION:
; APPLICANT: WEINER, RONALD M.
; TITLE OF INVENTION: MARINE MELA GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WATSON COLE STEVENS DAVIS, P.L.L.C.
; STREET: 1400 K STREET NW
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-2477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,254
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PODLOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: JAP30319C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-0088
; TELEFAX: 202 628-8034
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-476-254-7

Query Match 14.6% Score 155; DB 2; Length 153;
Best Local Similarity 29.7%; Pred. No. 1.2e-10;
Matches 38; Conservative 21; Mismatches 51; Indels 18; Gaps 2;

OY 57 WRGCTLFSSRVIA-----FVYKMGAAVYIMIGTOORAGAI-DL 98
DB 18 WMWGEELGVALVAIAAVMGVASMILNYPOLFIDILKMWGGLVIGTIGISMRAKGRKMANL 77
OY 99 KSLASTOSRRHLFORAVFVNLTNPKSIVFLAALFPOFIMPQOPOLMQYIVLGVTIVVDI 158
```

Db 78 DNTSSQSNKALITGCGVTAIANPKGMAMISLLPFTISVQALAPQMLVLLSIIMTEF 137

QY 159 IVMIGVAT 166
Db 138 FSLMAYAS 145

RESULT 2

5474933-4
; Patent No. 5474933
; APPLICANT: EINER, RONALD M.; FOUOA JR., WILLIAM C.
; TITLE OF INVENTION: MARINE MEIA GENE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,945
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 974,837
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: 496,804
; FILING DATE: 21-MAR-1990
; SEQ ID NO: 4:
; LENGTH: 153
5474933-4

Query Match 14.6%; Score 155; DB 6; Length 153;

Best Local Similarity 29.7%; Pred. No. 1,2e-10;

Matches 38; Conservative 21; Mismatches 51; Indels 18; Gaps 2;

QY 57 WRGVGLFSRSVIA-----FEVYKMGAGAYLIMLGIOQWRAGAL-DL 98
Db 18 WAAVAGELAGVAAVATAAVGVAASMLNTPQLPDLIKWVGLTIGTISMRAGKMANL 77
QY 99 KSLASTQSRRLHFORAVFVNLTNPKSIVFLAALFPOFTMPQPOLMOTIVLGVTTIVVDI 158
Db 78 DNTSSQSNKALITGCGVTAIANPKGMAMISLLPFTISVQALAPQMLVLLSIIMTEF 137
QY 159 IVMIGVAT 166
Db 138 FSLMAYAS 145

RESULT 3

US-09-068-195-24
; Sequence 24, Application US/09068195B
; Patent No. 6140078
; GENERAL INFORMATION:
; APPLICANT: Sanders, Jan W.
; APPLICANT: Ledebner, Adrianus M.
; APPLICANT: Venema, Gerard
; TITLE OF INVENTION: Salt-inducible Promoter Derivable from a Lactic Acid
; TITLE OF INVENTION: Bacterium, and Its Use in a Lactic Acid Bacterium for
; TITLE OF INVENTION: Production of Desired Protein
; FILE REFERENCE: Sanders-60113/0252227
; CURRENT APPLICATION NUMBER: US/09/068,195B
; CURRENT FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: PCT/EP97/04755
; EARLIER FILING DATE: 1997-08-20
; EARLIER APPLICATION NUMBER: EP 97200744/7
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: EP 96202444/4
; EARLIER FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 503
; TYPE: PRT
; ORGANISM: L. lactis MG1363
US-09-068-195-24

Query Match 6.8%; Score 72; DB 4; Length 503;
Best Local Similarity 21.1%; Pred. No. 3.6;
Matches 41; Conservative 25; Mismatches 66; Indels 62; Gaps 10;

QY 9 YLITSLITLSPGSGAINTMTTSLNHGYPAGVYCMASDRGDCYCAGMVGTLSR-- 66
Db 38 FLLGLGLMFLPVALCAEMAMTV--EGMKNGITFSWTSQIGERF-----GFAAIFPQWF 90
QY 67 -----SVIAFEVL-----KWAGAAVLIWLGIOQWRAGALDKSLA 102
Db 91 QITVGVMTIYFLLGALSYVNFQALNTPDLIKFGLIIFW-----GLTFSQLG 140
QY 103 STQSRRLHFORAVFVNLTNPKSIVF-LAALFPOFT-----POLMOTIVLG 150
Db 141 GQGRAKLVKAGFVYGIYIPSVILFGLAAAL--FIGGNPIEIPINSHAFVPSQ----- 193
QY 151 VTTIVVDIIVMIGY 164
Db 194 VSTLVFVFSFILAY 207

RESULT 4

5432081-9
; Patent No. 5432081
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
; GLUCONIDE PERMEASE GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,546
; FILING DATE: 15-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 447,976
; FILING DATE: 08-DEC-1989
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; SEQ ID NO: 9:
; LENGTH: 428
5432081-9

Query Match 6.6%; Score 70; DB 6; Length 428;
Best Local Similarity 26.1%; Pred. No. 4.9;
Matches 58; Conservative 27; Mismatches 101; Indels 36; Gaps 13;

QY 7 PAYLITSLITLSPGSGAINTMTTSLNHGYPAGVYCMASD--RTGDSYCA-GWR----- 58
Db 14 FAFAGALFL-LSYTYDVAGVGAARAHAVYTGAGIRCLRRLRTSGGQCEYRWGRFRFP 72
QY 59 --GVGTLSRSVIAFEVL-KWAGA-----AYLIMLGIOQWRAGALDKSLAST----- 104
Db 73 LLEFAPIMTISVLFVWLTDMSHGSKVYAYLTMGLGICYSLVNIPGSLATANTQOP 132
QY 105 OSRRHL-FORAVFVNLTNPKSIVFLAALFPOFTMPQPOLMOTIVLGVTTIVVDIIVMIG 163
Db 133 OSRARLGAARGLASLT---FVCLAFLLIPSPKSNSSPEEM-VSVYHFXITVLAIAAGVNL 187
QY 164 Y----ATLAORIALIMIKPKOMKLINKIFS--LEML-VGAL 198
Db 168 YFICKKSTRENVAVITVAOPSLNISLQTLKNNRPLFMCLGAL 229

RESULT 5

US-08-875-223-8
; Sequence 8, Application US/08875223
; Patent No. 6127175
; GENERAL INFORMATION:
; APPLICANT: VIGNE, Emmanuelle
; APPLICANT: PERRICAUDET, Michel
; APPLICANT: DEDIEU, Jean-Francois

Mon May 7 11:08:02 2001

us-09-466-935-2.raii

Page 3

[illegible]

```

1      NUMBER OF SEQUENCES: 33
2      CORRESPONDENCE ADDRESS:
3      ADDRESSEE: McCutchen, Doyle, Brown & Enersen, LLP
4      STREET: 3 Embarcadero Center
5      CITY: San Francisco
6      STATE: CA
7      COUNTRY: USA
8
9      ZIP: 94111
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE: Diskette
12     COMPUTER: IBM Compatible
13     OPERATING SYSTEM: DOS
14     SOFTWARE: FASTSEQ for Windows Version 2.0
15
16     CURRENT APPLICATION DATA:
17     APPLICATION NUMBER: US/08/448,603A
18     FILING DATE: 07-JUN-1994
19     CLASSIFICATION: 435
20
21     PRIOR APPLICATION DATA:
22     APPLICATION NUMBER: 08/072,833
23     FILING DATE: 07-JUN-93
24
25     ATTORNEY/AGENT INFORMATION:
26     NAME: Haliday, Emily
27     REGISTRATION NUMBER: 38903
28     REFERENCE/DOCKET NUMBER: 14918-704
29
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: 415-393-2000
32     TELEFAX: 415-393-2286
33
34     TELEX:
35
36     INFORMATION FOR SEQ ID NO: 30:
37
38     SEQUENCE CHARACTERISTICS:
39
40     LENGTH: 857 amino acids
41     TYPE: amino acid
42     STRANDEDNESS: single
43     TOPOLOGY: linear
44
45     US-08-448-603A-30
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```


Db 844 IFPVCSITGGLILAHVAMKIAVLL--AAPPVMITAGYVRRLVLAESRRSAVNDAA 902
 QY 116 EVNLTKNSIVFLAALFPQFIMPODLMQYVLT-----TIYVDI 158
 Db 903 SIAEACRGIRTIASIGRE-----RGVSRASNAVPEPYDKGIRFLITNT 948
 QY 159 IVMIGYAT--LAORIALMWIKPKOMKALNKIFGSL--FMVGLALLASARHA 205
 Db 949 LIALSEITFVYALAW--WGAQQR--NGTYSQDLEFIVLALLFSQAOSA 996

RESULT 10
 US-08-358-171-2
 ; Sequence 2, Application US/08358171
 ; Patent No. 5763578
 ; GENERAL INFORMATION:
 ; APPLICANT: FONG, Henry K.W.
 ; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
 ; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/358,171
 ; FILING DATE: 16-DEC-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: YUN, Allen C.
 ; REGISTRATION NUMBER: 37,971
 ; REFERENCE/DOCKET NUMBER: FONG=2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ. ID NO.: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 291 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-358-171-2

Query Match 6.3%; Score 66.5; DB 1; Length 291;
 Best Local Similarity 21.7%; Pred. No. 7.4;
 Matches 51; Conservative 35; Mismatches 76; Indels 73; Gaps 13;
 QY 11 LTSILTLSPGSGAINM--TSLNHGYPAGGYCWMASDRGDSYCAWMRG---VGNL 63
 Db 53 LVIISLALADSGISLNLVAATSSLLRNPYSGEQAH-----GFGFVYALASI 103
 QY 64 FSRSVIAF-----EVLKWAQA--AYLWLGIQWRRA-----GAILDKSLAS--- 103
 Db 104 CSSAAVAMGRHYHFCRSLDMNTAVSLVFWLSAFAWALPLLGWGHYDEPLEGTCT 163
 QY 104 ---TQSRRLHFGRAVYVNLNPKSIYFLAALFPQFIMPODLMQYIVG-----VTI 154
 Db 164 LDYSRGDRN-FTSFLF-----TMAFENFLDPLFTTVSYRLMDQ-KLGKTSRPVNTV 214
 QY 155 VVDIIVWIG-----YATLAQRI-----ALWIKPKOMKALNKIFGS 190
 Db 215 LPARTLLGWGPYALLVYIYATIDATSI SPKLQWVPALIAKAVPTVNMANNVALGS 269

RESULT 11
 US-09-090-947-2
 ; Sequence 2, Application US/09090947
 ; Patent No. 6008338
 ; GENERAL INFORMATION:
 ; APPLICANT: FONG, Henry K.W.
 ; TITLE OF INVENTION: ALL TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
 ; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/090,947
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/358,171
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: YUN, Allen C.
 ; REGISTRATION NUMBER: 37,971
 ; REFERENCE/DOCKET NUMBER: FONG=2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ. ID NO.: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 291 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-090-947-2

Query Match 6.3%; Score 66.5; DB 3; Length 291;
 Best Local Similarity 21.7%; Pred. No. 7.4;
 Matches 51; Conservative 35; Mismatches 76; Indels 73; Gaps 13;
 QY 11 LTSILTLSPGSGAINM--TSLNHGYPAGGYCWMASDRGDSYCAWMRG---VGNL 63
 Db 53 LVIISLALADSGISLNLVAATSSLLRNPYSGEQAH-----GFGFVYALASI 103
 QY 64 FSRSVIAF-----EVLKWAQA--AYLWLGIQWRRA-----GAILDKSLAS--- 103
 Db 104 CSSAAVAMGRHYHFCRSLDMNTAVSLVFWLSAFAWALPLLGWGHYDEPLEGTCT 163
 QY 104 ---TQSRRLHFGRAVYVNLNPKSIYFLAALFPQFIMPODLMQYIVG-----VTI 154
 Db 164 LDYSRGDRN-FTSFLF-----TMAFENFLDPLFTTVSYRLMDQ-KLGKTSRPVNTV 214
 QY 155 VVDIIVWIG-----YATLAQRI-----ALWIKPKOMKALNKIFGS 190
 Db 215 LPARTLLGWGPYALLVYIYATIDATSI SPKLQWVPALIAKAVPTVNMANNVALGS 269
 RESULT 12
 US-08-468-812-2
 ; Sequence 2, Application US/08468812
 ; Patent No. 5935836
 ; GENERAL INFORMATION:
 ; APPLICANT: Vehmaanper, Jari

```

; APPLICANT: M ntyl, Arja
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Iantto, Raija
; APPLICANT: Paloheimo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lahtinen, Tarja
; APPLICANT: Kristo, Paula
; TITLE OF INVENTION: Actinomadura Xylanase Sequences and Methods
; TITLE OF INVENTION: of use
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLSTEIN & FOX
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,812
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/332,412
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/282,001
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Larry B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 1050.0340002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-812-2

Query Match      6.3%; Score 66.5; DB 2; Length 344;
Best Local Similarity 22.7%; Pred. No. 9.3;
Matches 25; Conservative 11; Mismatches 45; Indels 29; Gaps 3;

QY 10 LLSIILTLSPGSGAINTMTSLNIGYPAGVYCWASD-----RTGSDYCGARGV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28 LALAIGALLPGTAHADTTITONOTGNDNGYRYSFTTAPGTVSKTLHSGGSYSTSWRNT 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 GTLFS-----RSVIAFEVLKWAAGAYLWLG-----IQQW 90
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 GNFVAGKGMSTGGRTVTYNASFNPSGNGYLTLYGWTNPLVEYYIYVSW 137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-08-863-102-1
; Sequence 1, Application US/08863102
; Patent No. 6015828
; GENERAL INFORMATION:
; APPLICANT: Cuppoletti, John
; TITLE OF INVENTION: Chemical Modification of
; TITLE OF INVENTION: Chloride Channels as a Treatment for Cystic
; TITLE OF INVENTION: Fibrosis and Other Diseases
; NUMBER OF SEQUENCES: 4
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jackie A. Zurcher,
; ADDRESSEE: Dinmore & Shohl LLP
; STREET: 255 E. Fifth Street,
; STREET: 1900 Chemed Center
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,102
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zurcher, J. A.
; REGISTRATION NUMBER: 42,251
; REFERENCE/DOCKET NUMBER: 10738-12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 977-8377
; TELEFAX: (513) 977-8141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rabbit Gastric CIC-2
; US-08-863-102-1

Query Match      6.3%; Score 66.5; DB 3; Length 898;
Best Local Similarity 25.0%; Pred. No. 35;
Matches 47; Conservative 26; Mismatches 76; Indels 39; Gaps 10;

QY 39 GGVYCW---ASDRT-----GDSYCGARGVGRFLFRSVIAFEVLKWAAGAYLWLGIO 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 GGEPEWRSPSPRPPELLEFGSBCARCMGCVRKHFVSRGDEW---ITVLGLGL 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 QWRAGAIDLKSLASTOSRHLFORAVFVNL-----TNPSIYFLAALFPQFIIMPQO 140
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104 MALVSAMDYALAACLOAQOMW-SRGINTNLLQYLAWVTYPPVLTFSAGFTQILAPQA 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 141 -----PQIMQYIVIGVT-----TIVVDIYWIGY-ATLAQRILMWIKP-----KQMKAL 184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 VQSGIPE-MKTLRGVVLKEYLTTLTKVAKYIGLTALGSGMPLKGGPVTHTASMCAL 221
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 NKIFGSLF 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 LSKFLSLF 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-863-102-4
; Sequence 4, Application US/08863102
; Patent No. 6015828
; GENERAL INFORMATION:
; APPLICANT: Cuppoletti, John
; TITLE OF INVENTION: Chemical Modification of
; TITLE OF INVENTION: Chloride Channels as a Treatment for Cystic
; TITLE OF INVENTION: Fibrosis and Other Diseases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jackie A. Zurcher,
; ADDRESSEE: Dinmore & Shohl LLP
; STREET: 255 E. Fifth Street,
; STREET: 1900 Chemed Center
; CITY: Cincinnati
; STATE: Ohio
```

```

? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/802,141
? FILING DATE: 19-FEB-1997
? CLASSIFICATION: 424
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 08/231,041
? FILING DATE: 15-APR-1994
? ATTORNEY/AGENT INFORMATION:
? NAME: Spratt, Gwendolyn D.
? REGISTRATION NUMBER: 36,016
? REFERENCE/DOCKET NUMBER: 141A, 609
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 404/688-0770
? TELEFAX: 688-9880
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 772 amino acids
? TYPE: amino acid
? TOPOLOGY: linear
? MOLECULE TYPE: protein
US-08-802-141-4

Query Match      6.2%; Score 66; DB 1; Length 772;
Best Local Similarity 20.0%; Pred. No. 33;
Matches    38; Conservative   28; Indels    66; Gaps      8
```

```

Query Match: 6.2%; Score 66; DB 1; Length 772;
Best Local Similarity 20.0%; Pred. No. 33;
Matches 38; Conservative 28; Mismatches 58; Indels 66; Gaps

QY      5 WMEAYLLTSLITLLSPGSGAINTMTTSLNHGYPAGGYCMASDRTGDSYCAGMNGVETLF 64
      |||  ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      102 WMETYL--VLPPTAQTSTVYVVMNETVN-----ISIDNSGSAV      137

QY      65 SRSTIAEVLKMGAGAVLLIMLGQMRAGATIDTLKSLASTOSRHLEFORAVFVNLTN-- 121
      |||  ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      138 -KFDVDEKT-----SSQAQAKSRNRFLLTAHRLQAYRRDDGMSINVMG 179

QY      122 -----PKSIVFLAALFPQFINP--QQPQLMOYIVLVGTTI--VVDIIVMI 162
      |||  ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      180 SDAYGDLRGVTFENPVNVAVINLNADF--YVLPDSQOEKCTEYIRRGPAITOTTTVYTP 237

QY      163 GYATLAQRIA 172
      ||  ::  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      238 SYAVRSORIA 247

Search completed: May 6, 2001, 14:39:23
Job time: 414 sec

```

Search completed: May 6, 2001, 14:39:23
Job time: 414 sec

OM of: US-09-466-935-2 to: GenEmbl:* out_format : pfs
Date: May 6, 2001 3:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-O=/gen2.1/USPTO_spool/US09466935/runat_03052001.075709_14798/app_query.fasta_1.529
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMAPCH=0.100 -LOCPCL=0.000 -LOEPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FCGAPOP=6.000
-FCGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELCP=6.000
-DELPEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.ccd1
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09466935.@CGN1.1 5396 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-466-935-2
Query Length: 205
Database: GenEmbl:*
Database sequences: 1283235
Database length: -1216004940
Search time (sec): 1745.950000

Score list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat1:AX030175	+	1061.00	1875.45	3.0e-96	1231	AX030175 Sequence 1 from Paten
gb_ba2:ECOM85	+	1039.00	1796.54	7.3e-92	91414	M87049 E. coli genomic sequen
gb_ba1:AE000458	-	1024.00	1788.94	2.0e-91	11509	AE000458 Escherichia coli K12
gb_ba1:AE0005613	-	950.00	1657.93	3.9e-84	10592	AE0005613 Escherichia coli O15
gb_ba3:STSTMT1	-	845.00	1450.61	1.4e-72	96086	AF233324 Salmonella typhimur
gb_hlg7:AC020870	-	795.00	1352.52	4.0e-67	256373	AC020870 Mus musculus clone
gb_hlg7:AC022157	-	753.00	1311.69	7.6e-65	4883	AC022157 Homo sapiens chromosc
gb_hlg7:AC020970	-	734.00	1245.60	3.7e-61	212936	AC020970 Mus musculus clone
gb_hlg7:AC020874	+	637.00	1070.84	2.0e-51	265383	AC020874 Mus musculus clone
gb_hlg11:AC026742	-	477.00	790.02	8.7e-36	170178	AC026742 Homo sapiens chrom
gb_ba1:AE004104	-	460.50	789.21	9.6e-36	11305	AE004104 Vibrio cholerae chrc
gb_ba1:AE0004937	-	402.50	679.51	1.2e-29	10475	AE0004937 Pseudomonas aerugin
gb_hlg7:AC020885	+	397.50	644.25	1.1e-27	268294	AC020885 Mus musculus clone
gb_hlg7:AC020833	+	270.50	431.56	8.0e-16	62274	AC020833 Mus musculus clone
gb_ba2:AP003001	+	255.50	389.04	1.9e-13	34783	AP003001 Mesorhizobium loti
gb_ba2:AP003005	+	235.50	353.36	1.8e-11	347950	AP003005 Mesorhizobium loti
gb_ba2:AP003013	+	228.50	341.31	8.5e-11	332635	AP003013 Mesorhizobium loti
gb_ba1:AE0002037	-	228.00	370.87	1.9e-12	12268	AE0002037 Deinococcus radiodur
gb_ba2:AP000295	-	218.50	323.12	8.8e-10	346897	AP000295 Mesorhizobium loti
gb_ba1:AF157493	+	203.50	320.51	1.2e-09	25454	AF157493 Zymomonas mobilis ZM
gb_ba2:AP003012	-	200.50	291.07	5.4e-08	346547	AP003012 Mesorhizobium loti
gb_ba2:BSDB0014	-	197.00	289.31	6.7e-08	213420	Z59117 Bacillus subtilis com
gb_pat1:AX030081	+	195.50	340.25	9.8e-11	639	AX030081 Sequence 11 from Paten
gb_ba1:AE000274	-	195.50	311.91	3.7e-09	13793	AE000274 Escherichia coli K12
gb_ba2:DP90824	-	195.50	311.62	3.9e-09	14236	DP90824 E. coli genomic DNA, KC
gb_ba2:DP90823	-	195.50	310.85	4.3e-09	13475	D90823 E. coli genomic DNA, KC
gb_ba1:AE016260	-	195.50	286.95	9.1e-08	206479	AE016260 Agrobacterium tumef
gb_ba1:AE000484	-	193.00	309.57	5.0e-09	10976	AE000484 Pseudomonas aerugin
gb_ba2:AP003008	+	193.00	277.71	3.0e-07	346820	AP003008 Mesorhizobium loti
gb_ba2:AP001568	+	190.50	274.66	4.4e-07	296950	AP001568 Bacillus halodurans
gb_ba1:AE0004221	-	189.00	303.87	1.0e-08	11401	AE0004221 Vibrio cholerae chrc
gb_ba2:AP003004	-	189.50	271.70	6.4e-07	338579	AP003004 Mesorhizobium loti
gb_ba2:AP003012	+	187.50	297.93	1.0e-06	346547	AP003012 Mesorhizobium loti
gb_ba1:AE0004372	-	186.50	298.64	2.0e-08	10176	AE0004372 Vibrio cholerae chrc
gb_ba1:AE0004786	-	186.50	298.64	2.0e-08	10235	AE0004786 Pseudomonas aerugin
gb_ba1:AE0005402	-	186.50	297.71	2.3e-08	11313	AE0005402 Escherichia coli O15
gb_ba1:AB024601	-	186.50	295.23	3.2e-08	14807	AB024601 Pseudomonas aerugin
gb_hlg75:LMFICHR32_20	+	186.00	278.84	3.8e-07	110000	Continuation (21 of 28) of
gb_pat1:AR061372	+	184.00	316.77	2.0e-09	885	X67020 S. colwelliana mlgA gene
gb_ba3:SCMIGA	+	184.00	316.77	2.0e-09	885	X67020 S. colwelliana mlgA gene
gb_ba1:AF188935	+	184.00	273.51	5.1e-07	96231	AF188935 Bacillus anthracis p

gb_ba1:AE004889	-	180.50	286.95	9.1e-08	11415	AE004889 Pseudomonas aerug
gb_ba1:AE004079	+	179.00	285.09	1.2e-07	10445	AE004079 xyliella fastidios
gb_ba3:PACARAB	+	178.50	297.46	2.4e-08	2483	U04992 Pseudomonas aerugin
gb_ba3:PAU81259	+	178.50	287.53	8.5e-08	7285	U81259 Pseudomonas aerugin

seq_name: gb_pat1:AX030175

seq_documentation_block:

LOCUS AX030175 1231 bp DNA PAT 16-SEP-2000
DEFINITION Sequence 1 from Patent EP1013765.
ACCESSION AX030175
VERSION AX030175.1 GI:10190392
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 1231)
Belareva, A.V., Aleashin, V.V., Livshits, V.A., Tokmakova, I.L. and

AUTHORS
Zakateeva, N.P.
Gene and method for producing l-amino acids
Patent: EP 1013765-A 1 28-JUN-2000;
AJINOMOTO KK (JP)

FEATURES
location/Qualifiers

source 1..1231
/organism="Escherichia coli"
/db_xref="taxon:562"
57..1174
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09318.1"
/db_xref="GI:10190393"
/translation="MILEWFAVLITSLTSPGSAINTMTTSLNHGYPAGGYCM
ASDRTGSGCAGMGVGTLSRSYIAEVEIKMGAAFLIMGIDQWRAAGADIDKSLA
STOSRHLFORAVVNLTPKRSIVFLALPPQIMPQOLMOYIYGVYTIYVDITV
MIGVATLAORIALIMIKGPKMKMLNKFESLFLMVGALLASARHA"

CDS
BASE COUNT 291 a 319 c 311 g 310 t
ORIGIN

alignment_scores:
Quality: 1061.00 Length: 205
Ratio: 5.176 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-466-935-2 x AX030175 ..
Align seg 1/1 to: AX030175 from: 1 to: 1231

1	MetThrLeuGluTrpTrpPheAlaTrpLeuLeuThrSerIleIleLeuTh	17
557	ATGACCTTGAATAGTGGTTGGCTTGCATCTGACATCATCATTTTAA	606
17	rleuSerTrpGlySerClyAlaIleAsnThrMetThrThrSerLeuAsn	34
607	GGCTGCGCAGGCTGTGGCAATCAACATGACACCTGCTGCACAC	656
34	isGlyTrpProAlaGlyGlyValTYrCysTrpAlaSerAspArgThrGly	50
657	ACGGTATTCGGCGCGGTGCTGTATTCGTGGCTTACACCGGACTGCG	706
51	AspSerTYrCysAlaGlyTrpArgGlyValGlyThrLeuPheSerArgSe	67
707	GATTCATATTTGGTGGTGGCTGGGCTGGGACGATTTTCCCGCTC	756
67	rValIleAlaPheGlyValLeuLeuTrpAlaGlyAlaIleAlaTYrLeuIle	84
757	AGTGATTTGGCTTGAAGTGTGAAGTGGCAGCGCGGCTTACTGATT	806
84	rLeuGlyIleGlnGlnTrpArgAlaIleAlaIleAspLeuYsser	100

807 GGCTGGGAATCCACGACGATGGCGCCGCTGGTGCATTCACCTTAATCG 856
101 leuA1aserThrgInserArgArqHisLeuPheGlnarGalalValPheVa 117
|||||
857 GTGGCCCTCTACTCAATCGCGTCGACATTTGTTCCAGCGCCAGTTTGT 906
117 lAsnLeuThAsnProlySserileValPheLeuAlAlaleuPheProG 134
|||||
907 GAATCTCACCACATCCCAAAAGTATTGTTTCGGCGCGCTATTTCGCG 956
134 lnpHeileMetProGInGlnProGInleuMetGlnTrrileValleuGly 150
957 AATTCATCATCGTCGCAACAGCCGCACTGCATGCATTCGCTGCGGC 1006
151 ValThrThrIleValAlaAspIleIleValMetIleGlyTrrAlaThrle 167
|||||
1007 GTACCACTATTGGTCGATATTATTGTGATGATCGTTACGCCACCT 1056
167 uAlaGlnArgIleAlaLeuTrrileLysGlyProLysGlnMetLysAlaL 184
|||||
1057 TGCACAACGAGATTGCTCTATGATTAAGACCAAGCAGATGAAGCGCG 1106
184 eusnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuAla 200
|||||
1107 TGATATAGATTTCGCTGCTGTTATGCTGTGGAGCGCTGTATGCA 1156
201 SerAlaArgHisAla 205
|||||
1157 TCGCGAGGCGCATGCG 1171

seq_name: gp_ba2:ECOW85

seq_documentation block:

LOCUS ECOW85 91414 bp DNA BCT 29-MAY-1995

DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.

ACCESSION M87049

VERSION M87049.1 GI:836656

KEYWORDS

SOURCE

ORGANISM

Escherichia coli.

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

Escherichia.

1 (bases 1 to 91414)

Daniels,D.L., Plunkett,G. III., Burland,V. and Blattner,F.R.

Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes

Science 257 (5071), 771-778 (1992)

2 (bases 1 to 91414)

Plunkett,G. III., Burland,V., Daniels,D.L. and Blattner,F.R.

Analysis of the Escherichia coli genome. III. DNA sequence of the region from 87.2 to 89.2 minutes

Nucleic Acids Res. 21 (15), 3391-3398 (1993)

3 (bases 1 to 91414)

Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and Daniels,D.L.

Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes

Nucleic Acids Res. 21 (23), 5408-5417 (1993)

4 (bases 1 to 91414)

Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and Rouviere,P.E.

A new family of peptidyl-prolyl isomerases

Trends Biochem. Sci. 20 (1), 12-14 (1995)

5 (bases 1 to 91414)

Daniels,D.L.

Direct Submission

Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics,

University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecolli@genetics.wisc.edu; Phone: 608-262-2534; Fax:

COMMENT

608-263-7459
On May 29, 1995 this sequence version replaced gi:148169.
This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the university of Wisconsin-Madison. The entire sequence was independently determined from E. coli Mgl665. Overlaps and conflicts with other sequence determinations are annotated. Reference [1] describes the original sequence determination of a 91408 bp sequence. References [2], [3], and [4] describe subsequent corrections and/or updates to that sequence.

FEATURES

source

1..91414
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/clone_lib="lambda library of Daniels and Blattner"
/map="bp 1 at 3974 kb; 84.7 minutes"
/note="This sequence comprises the following lambda clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177), DD856(EC19RM8.1), DD858(EC17RM9), DD861(EC17-221), DD864(EC27-149), DD865(EC21-76), DD867(EC27-860). M13mp19 or Janus vectors were used for subcloning."

misc_feature

1..200
/note="corresponds to K02846 (1..200)"

misc_feature

1..5863
/note="clone EC15-116"
complement(1..4899)
/note="corresponds to M37337 (1..4900)"

misc_feature

1..339
/note="corresponds to V00349 (164..500)"

tRNA

10..86
/gene="aspT"

/note="CGSC No. 989"

gene

10..86
/gene="aspT"

95..170
/gene="trpT"

tRNA

95..170
/note="CGSC No. 66"

95..170
/gene="trpT"

gene

177..205
/gene="trpT"

177..205
/note="terminator-like sequence"

terminator

253..396
/note="predicted bend of 81 degrees"

misc_structure

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

gene

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

CDS

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

misc_structure

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

misc_structure

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

gene

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

CDS

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

misc_structure

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

gene

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

CDS

complement(266..862)
/gene="f198"

complement(266..862)
/gene="f198"

promoter complement(1136. .1165)
/note="promoter matrix score of 63"
1151. .1179
promoter /note="promoter matrix score of 57"
1223. .1636
gene /gene="0137"
1223. .1636
CDS /gene="0137"
/codon_start=1
/transl_table=1
/function="unknown"
/protein_id="AA067568.1"
/db_xref="GI:148172"
/translation="MAESFTTNKYPDNKHYPGRGSRGDTIKEADLREHCAENE
LDGKREPVTSEKLEVAVCGEREPEVTAEARVSKYWTIRKPRFHTLSAVNRRLK
VLKTPPIITINKKEMPFFYVSCFRMMSKRSIAR"
1562. .1586
terminator /gene="0137"
/note="terminator-like sequence"
complement(1585. .3135)
/gene="F516"
CDS complement(1585. .3135)
/gene="F516"
/note="similar to Mg chelatase subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=1
/function="unknown"
/protein_id="AA067569.1"
/db_xref="GI:148173"
/translation="MHIMVTOGGLMSLSIVHTRAALGNAPITVEVHSIKGLIIM
VSLPRTYKARDRYRSALINSGYEYPAKTTITLAPRDLPEGGRDIPFAITLLA
SBLTANKUDEIYELGELATLGALKEVPGLSSATEATKSGRKITIVADNDEGLIN
GGCIIADHQAVCAPLEKHALEKPRPDVASRALQDLSDVIGEGKGLDITPA
GGHNLILGPQGTMTLASRINGLIPDLSBEALLESAILSYNAESVQKQMPORF
RSPHSASLTAAVGGGAIIPGEISLAHNGVLFDELPEFRRTLDALREPIESGQIL
LSRTAKITTPARFQIVAAIMPSPIGHYOGHNHCTPOTLRYLNRISGPELDLDS
LEIIPPGIILSKTYVPGSSATYVQRVMAAREQFRKONLNMVDSPETRQCKLE
SEDAWMLBGTLLHGLSTRAWORLKVARTIADIDDSIITRHLQEAHSVIRAIKRL
IHLQKILT"
3287. .3433
misc_structure /note="predicted bend of 79 degrees"
3297. .3326
promoter /note="promoter matrix score of 74"
3336. .3676
misc_feature /note="corresponds to M11651 (1. .341)"
3336. .3737
misc_feature /note="corresponds to V00289 (1. .402)"
3336. .12791
misc_feature /note="corresponds to M10313 (1. .9456)"
3379. .5721
gene /note="V00290; ECILVX(1. .2343)"
3458. .3556
gene /gene="11VL"
3458. .3556
CDS /gene="11VL"
/codon_start=1
/transl_table=1
/product="l1vGMDA operon leader peptide"
/protein_id="AA067570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVVISVVIIPCCAGALGRGKA"
3696. .4679
gene /gene="11VG"
3696. .4679
CDS /gene="11VG"
/note="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=1
/product="acetohydroxy acid synthase II, large subunit"
/protein_id="AA067571.1"
/db_xref="GI:148175"
/translation="MNGAOMVYVHALRAGCVNTFVGPAGAIIMPYDALYDGGVEHLIC

RHOGAAMAIGARATGKTGVCATSGPATGNTLTGLADALIDSIPVATIGVYAP
FICTDAFOEVDVIGSLACTKHSFIVQSLLEELPRIMAEAFDVAGSRGPPVVDIPKD
IQLASDLEBWFTTVENEVYFPAEVEQRAROKRPMILYVGGYGAQAVPRLRE
FLAATMPATCTLKGAAVEADPYTLGMLGHMGTAAVQEBCLIAVGARDDK
VTGKLNTPAPHASVITHMDIDPAEMNKLROAHVALQGDINALPALQOPLNO"
4675. .5340
gene /gene="0221"
4675. .5340
CDS /gene="0221"
/note="the mutation l1v02096 joins this ORF to l1v6,
restoring function missing in wild-type"
/pseudo
/codon_start=1
replace(4677. .4678, "tatg")
variation
alignment_scores:
Quality: 1039.00 Length: 206
Ratio: 5.093 Gaps: 1
Percent Similarity: 99.029 Percent Identity: 99.029
alignment_block:
US-09-466-935-2 x ECOUW85/rev ..
Align seg 1/1 to reverse of: ECOUW85 from: 1 to: 91414
1 MetThrLeuGluTrpTrpPheLeuThrLeuLeuThrSerIleLeuLeu 17
|||||
62161 ATGACCTTAGAATGGTGGTGGCTTACCTGCGACATGATCATTTTAAAC 62112
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
|||||
62111 GGTGTCGCCAGGCTCTGGTGCACATACACACTATAGCCACCTGCGCAACC 62062
34 IsgLYTyr. ProIaGlyGlyValIlyrcysTrpAlaSerAspArgThcI 50
|||||
62061 ACGGTATACSCGCGGCGGCGCTATTGCTGGGCTTACAGACCGGACCTGG 62012
50 yAspSerTyrcysAlaGlyTrpArgGlyValGlyThrLeuPheSerArgS 67
|||||
62011 CATTTCATATTGTGCTGGTGGCGGGGTGGAGCGCATTTTTCGGCT 61962
67 erValIleIleAlaPheGluValLeuIlyStrPaIaGlyAlaAlaIlyrLeuIle 83
|||||
61961 CAGTGAATTCGTTTGAAGGTGGAAGTGGCGAGCGGCGCTTACTTGAAT 61912
84 TrpLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeuIlySse 100
|||||
61911 TGGCTGGGAATCCAGCAGTGGCGCGCGCTGCTGCAATTGACCTTAAATC 61862
100 rLeuAlaSerThrGlnSerArgArgHisLeuPheGlnrArgAlaValrPheY 117
|||||
61861 GCTGGCCTTACTTCAATCGCTGCACATTGTTCACAGCGCGGAGTTTGTG 61812
117 aIAsnLeuThrAsnProIySerIleValrPheLeuAlaIleLeuPhePro 133
|||||
61811 TGAATTCACCAATCCCAAAAGTATTTGTTCTGGCGGCGCTATTTCCG 61762
134 GlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrlIleValLeuS 150
|||||
61761 CAATTCATCATGCGCGCAACCGCAACTGATGATATATGCTGCTGGG 61712
150 yValrThrTrpIleValrIleAspIleIleValMetIleGlyTrpAlaThrL 167
|||||
61711 CGTCACACACTATTGTGTGATATTATGTGATGATGGTTAGCGCACCC 61662
167 euAlaGlnArgIleAlaLeuTrpIleGlyProIySglnMetIlySala 183
|||||
61661 TTGCTAACGAGATTGCTTATGATTTAAAGGACCAAGAGAGTGAAGCGG 61612
184 LeuAsnIlyIlePheGlySerLeuPheMetLeuValGlyAlaIleuLeuAl 200
|||||
61611 CTGAATTAAGATTTTGCGCTGTTGTTATGCTGGTGGAGCGCCTGTATAC 61562

```

200 aSerAlaArgHisAla 205
|||||
61561 ATGGCGAGCGAATGCG 61546

seq_name: gb_ba1:AE000458

seq_documentation_block:
LOCUS      AE000458      11509 bp      DNA      BCT      01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete
            genome.
ACCESSION  AE000458 U00096
VERSION    AE000458.1 GI:2367299
KEYWORDS
SOURCE
ORGANISM   Escherichia coli K12.
            Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 11509)
            Blatter,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J.J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
97426617
PUBMED
REFERENCE  2 (bases 1 to 11509)
            Blatter,F.R.
            Direct Submission
            Submitted (15-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 11509)
            Blatter,F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 11509)
            Plunkett,G. III
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            On Sep 9, 1997 this sequence version replaced gi:1790254.
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.

```

FEATURES	
source	Location/Qualifiers
gene	<p>1..11509</p> <p>/organism="Escherichia coli K12"</p> <p>/strain="K12"</p> <p>/sub_strain="MG1655"</p> <p>/db_xref="taxon:83333"</p> <p>81..950</p> <p>/gene="pIdA"</p> <p>/note="b821"</p> <p>81..950</p>
CDS	<p>/gene="pIdA"</p> <p>/EC_number="3.1.1.32"</p> <p>/function="enzyme: Central intermediary metabolism: Phosphorus compounds"</p> <p>/note="O289; 99 pct identical to PAL_ECOLI SW: P00631; CGSNO..384"</p> <p>/codon_start=1</p> <p>/product="outer membrane phospholipase A"</p> <p>/protein_id="AAC76824.1"</p> <p>/db_xref="GI:2367300"</p>
promoter	<p>/translation="MRTLGQHLPEVFLPMANVYQGEATVEYHDAPEAVRGSITANMLQEHNDPELPYDVIYLYITQTSDLNKALIASYDAENARKDEVFQSLAFPLWRGILLGPNISVLASTYQKSMQLSNSSESPRENTYEPQLFAGATDPEFAGWTLIDVEMGYCINHNSGRSDPTSSMNRILYTRIMANGNMILVEVPMYVVGTDNDPDTIKMGYYQAKLIGYHGDPAVLISAKGOYVMNTGYGGAELGLSPITKHKVRLTYQVYSGSGESLIDVNFNQTRYGVGMLNDLF"</p> <p>965..993</p>
promoter	<p>/note="factor Sigma70; predicted +1 start at 4003392"</p> <p>1019..1047</p>
gene	<p>/note="factor Sigma70; predicted +1 start at 4003446"</p> <p>1077..2909</p> <p>/gene="recQ"</p> <p>/note="b3822"</p> <p>1077..2909</p> <p>/gene="recQ"</p> <p>/EC_number="3.6.1.-"</p> <p>/function="enzyme: DNA - replication, repair, restriction/modification"</p> <p>/note="o610; 99 pct identical to 607 amino acids of RECQ_ECOLI SW: P15043 but has 3 additional N-ter residues"</p>
CDS	<p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="ATP-dependent DNA helicase"</p> <p>/protein_id="AAC76825.1"</p> <p>/db_xref="GI:2367301"</p> <p>/translation="MNVAAQAEVLINLPSGAKOYLQETFGYQOPRPGOEILLIDYSGPDCLVYMPGGGKSLCYOIPALLINGLTVYVSPILSKDYOVDQANVCYAAACLNSTOTREQQLPMTGCRGQIRLTLIAPERLMLDNFLSLHAWNFVLAVDAHCISQGHDFRPFYALAGQLRQRPPLPFMALATADDTTODIIVRLGLNDPLIQISDFDRPIRYMLMKEFPLDLMKRYVOEORGSGLTIYCNSRAKVEDPAAQLSGSISAAAYHAGLENNVRAVQKRFQEDDQIIVATVAAFGMGINKPVRVYVHFDIPRNIESYQETFGAGRDGLPALAMLEFYPDADAMALRCLLEKRGQDGLDIEHHKINAMKAPFAEOTGCRBLVLIINFTGCGEAGCCGNDCTCLDPPKQYDSTDAQIALSTIGCNOPRCMGYVYVPTGANNQRRDYCHDILKTYGMRDKSKHEHWYVIRQLHILGIVTONIAOSHSLQLEAFKPVLAESSLQIALVRIVALKPKAMQKSGGYNVDRKLEFKRLKRSIADENSVPEYVNDATLIEMAEQMPITASEMILSVNGVGYMKRLDERGKPFMALIRAHVDDDEE"</p> <p>2104..2140</p>
repeat_region	<p>/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"</p> <p>3162..3189</p>
promoter	<p>/note="factor Sigma70; predicted +1 start at 4005588"</p> <p>3224..3592</p>
gene	<p>/gene="yigJ"</p> <p>/note="b3823"</p> <p>3224..3592</p> <p>/gene="yigJ"</p> <p>/function="orf; Unknown"</p> <p>/note="O122; sequence change shortens and changes N-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"</p> <p>/codon_start=1</p>
CDS	

[illegible]

AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Haybew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grothbeck,E.J., Davis,N.W., Lin,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.

TITLE Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

JOURNAL

FEATURES

source 1..10592
/organism="Escherichia coli O157:H7"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:83334"
/note="enterohemorrhagic"

misc_feature 1..71
/note="O-island #155; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
complement(72..836)
/gene="25332"
complement(72..836)
/gene="25332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to residues 1 to 131 of 161 from Escherichia coli K-12 Strain MG1655"
/codon_start=1
/transl_table=11
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKMITLNEKRIFLATLLPLFAVADDCALSDPTLTVOAYVNPOTERVKMYMOKANGAEAMGTTHALLAMNSOGVOMANGGITDESAPIGLYIEGQOKVALNLASGEENFEIRPGCVFAGKVCIVRIADAKTKEIQFAVQSGPMLMEMSVINPRRIHPVNASRKIRNGVGINKHGNAYFLISQATNIFYDFCAKAKLAVEQLLYDGTISIMYMGALIPWORYPVTMISVERKG"

gene 1206..2156
/gene="corA"
/note="25333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules: cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical to residues 1 to 316 of 316 from Escherichia coli K-12 strain MG1655: B3816"
/codon_start=1
/transl_table=11
/product="7g2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MLSAFOLENNRLTRLEVESOPLVNAVWIDIVEDDDDEFLRVS ELGOSIATREPELEDIASARPEEDDLHIHSPPFEDADHAGNSVATITDGRIFTIRRELPAPFLYRKARASQSMVDGNAYELLDFETKIQDLADEINITSIDLEQSRVIMSGHQDEYDEALSTLAELEDIGWVRLCLMDTORALNLFYKAKARLPGQGLEQARE LIRQIESLIPNESLFOKNVFLMQAMGFNIEQNRIKIKISYVSVVFLPPLVASSY GMEFEPMEIKMSFGYPGALIFMILAGLAPLYFKRKMWL"

misc_feature 2157..4848
/note="O-island #156; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="25334"
complement(2198..2689)
/gene="25334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MRSKLOKIIICLFLOCITNMTLRPVQVLLITYTSAGNSVFLVY DHEFWTSKDITNWLKHNELKNOHPPEBSGWTFVVIDGNSFTDYKRIEYEDLL YCFPTKSNDCITKNLYINEXYRNTHIGIDPTEYOLTDENKIEQVFNHNPKK

gene complement(2686..3534)
/gene="25335"
complement(2686..3534)
/gene="25335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNTLIPSSINIPCTIFETISLPDNYSADDMOYGDVDEDFLSIG LSDISAKVDYPRLYKXHPGPGSINVAFSSSGFTISQRECTDIFAFMKELAKFES FPGQVKTLEDLEHRRYNGSNFHSQQLNSPHEKINRYGNSPIRITKECIENGIN STPTGTOPLIDISIKTLSSRLNKFNDESPNGLGISVDHISAQKISLSFOKXA IGWSATVHFVADDFGLDVTDIKNKYSRREFRIFLQRRKDFAKPFPFNPNIE RIENYL"

gene 3748..3867
/gene="25336"
3748..3867
/gene="25336"
/function="orf; Unknown function"
/note="Residues 1 to 39 of 39 are 100.00 pct identical to residues 278 to 316 of 316 from GenPept 118 : g116960250|db|AAFF3440.1| (AF233324) S. typhimurium magnesium transport protein (CORA) (SP:P31138) [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="ANFEPMPLEKMSFGYPGALIFMILAGLAPLYFKRKMWL"

gene complement(3892..4278)
/gene="25337"
complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59013.1"
/db_xref="GI:12518691"
/translation="MEKSSIRNIFYPSPKLERLSRPROFIYTLTSMIITINDIWL SFLIYNNRYVININGAKRIPTGCKAIYFVLISGLHFLITISAIEGVRESLSVQ LVNAVILFTTILSAIPSAKMSLSAH"

gene 4483..4794
/gene="25339"
4483..4794
/gene="25339"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59014.1"
/db_xref="GI:12518692"
/translation="MDSITWMTYTPSSCSVMKNPKSKSDPYKAWLLMMAITIVLFI L SITTTUNLSIGISFGFCQTLGKRTMSLGSKIWGVGGFPVTNIVLFLILFFKR R"

gene complement(4841..5728)
/gene="rard"
complement(4841..5728)
/gene="rard"
/function="orf; Unknown function"
/note="Residues 4 to 295 of 295 are 92.12 pct identical to residues 4 to 299 of 300 from Escherichia coli K-12 strain MG1655: B3819"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59015.1"
/db_xref="GI:12518693"

```

/translation="MDAKOTRGVLLAIAAVFWIGTAPAVFKLIYYVADILLTRHVI
WSFFPMVLMSTICRQMSYKLTIOPTPKIPKPLAVSAVLIGCNMLFTIMNVNHHMLA
SLGFIPLVNIYVLSMPLGERFERMONLATILACVIVOLITRESPTILGLAES
PRTGLVKKINAEADOTGMLEFPMMLVPAITLFIADSDTHKONWASJNLILIA
AGIVTVPLCTFAAKTRLLSTLGFPIITPLMLAVTFGEKPGADKRVTTFTI
WVALAIFVMDAIIYTORRKS"
gene
complement(5780..6265)
/ gene="y1g1"
/ note="y5341"
/ complement(5780..6265)
CDS
complement(5780..6265)
/ gene="y1g1"
/ function="orf: unknown function"
/ note="Residues 1 to 161 of 161 are 100.00 pct identical
to residues 1 to 161 of 161 from Escherichia coli K-12
strain MG1655: B3820"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="A0359016.1"
/db_xref="GI:12518694"
/translation="MKCKDMSAVLTARQALRVGEMFVYHMPFRALGMELEERYEKE
PAQLAFKNQPMWVGNNAOSILHGVIALDVAAGLVCSFTLRHETISEDLQRL
SRMGIDLRVDYLRPGRXERETATSLLRACNKVAVARVELHDEOLYASATATYV
G"
gene
6412..7281
/ gene="p1da"
/ note="y5342"
6412..7281
CDS
/ gene="p1da"
/ function="enzyme: Central intermediary metabolism:
Phosphorus compounds"
/ note="Residues 1 to 289 of 289 are 100.00 pct identical
to residues 1 to 289 of 289 from Escherichia coli K-12
Strain MG1655: B3821"
/codon_start=1

alignment_scores:
Quality: 950.00 Length: 201
Ratio: 4.847 Gaps: 2
Percent Similarity: 97.512 Percent Identity: 96.517

alignment_block:
US-09-466-935-2 x AE005613/rev ..
Align seg 1/1 to reverse of: AE005613 from: 1 to: 10592

7 PhealatyrlleuLeuThrSerIleIleuThrIleuSerProglySerGI 23
|||||
10591 TTGGCTACTACTAGACATCAATCATTTTAACGCTGCGCAGCGCTCTGG 10542

23 valajleasphrmethrthrserleuasnhisgltyrprola..GI 39
|||||
10541 TGCATCAACACTGTGACACGCTGCTACACACGCTTATCGCGCGCGG 10492

39 ycllyaltlyrcystrrpalaseraspargthrglyaspserlyrcysalag 56
|||||
10491 TGGCGTCTATTGGCGGGCTTCACAGCGACTGCGATTCAATATTGTCTG 10442

56 lytrparaglyval.glythrleupheserargservallealaphegi 72
|||||
10441 GTTGGCGTGGGGTTGGGGAGCGATATTTCCGCTCAGATGGTGGTTGA 10392

72 uvalleuysrrtrrpaglyalaatyrlleuiletrpenclytlegeg 89
|||||
10391 AGTGTGAAGTGGGCGAGCGCGCTTACTGTGTTTGGCTGGGAATACAGC 10342

89 lntrrparaglyalaileaspleuysserleualaserthgin 105
|||||
10341 AATGGCGGCGCGTGGCGCAATGATCTAAATCGCTGCGCTACACAA 10292

106 serfargcgnhisleuphegninatrgalavalphevalasleuthrhanpr 122
|||||
10291 TCGCGTGCACATTTGTTCACAGCGGCAATTTTGTGAATCTCACCAATGCC 10242

```

```

122 olyserrlelevalpneleualaleupneproglinphellemetprog 139
|||||
10241 CAAAAGTATTGTTGTTCTGGCGCGCTATTTCGCGAATTCATCATACGCC 10192

139 lnginproginleugetcgtatrrlylevalleuglyvalthrthrlleval 155
|||||
10191 AACAGCGCAACTGATGATGATGATGCTGGGTGACACACTATTTGG 10142

156 valaspllellevalmetlleglytyrlealathrlleualacnarglleal 172
|||||
10141 GTCGATATTATGTGTGATGATGCTGATGCGCCACCTGCTCAAGCATTCG 10092

172 aleutrrpilleysglyprolysglmetlysalaleuasnllylephag 189
|||||
10091 TCTATGATTTMAAGACCAACACATGACGCGCTGATTAATAGATTTTGG 10042

189 lyserrleuphemelleualaglyalaleuLeuAlaserAlatrgHisAla 205
|||||
10041 GCTCGTTGTTATGCTGTGGAGCGCTGTTAGCATGGCGANGCATGGG 9992

seq_name: gb_ba3:STVSTMD1

seq_documentation_block:
LOCUS STVSTMD1 96086 bp DNA BCT 11-FEB-2000
DEFINITION Salmonella typhimurium fragment STMD1.
ACCESSION AF233324
VERSION AF233324.1 GI:6960215
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2.
Salmonella typhimurium LT2.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 96086)
AUTHORS Washington University Genome Sequencing Center.
TITLE The Salmonella typhimurium genome sequencing project
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 96086)
AUTHORS Watson, R.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: sc1ifton@watson.wustl.edu or
jspliehwatson.wustl.edu

NOTES:
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs
GLIMMER (Salzberg, S., Delcher, A., Kasif, A. and White, O. (1998)
NAR 26,544-548), and Genemark (Lukashin, A.V. and Borodovsky, M.
(1998), NAR 26,1107-1115.
FEATURES
source
1..96086
/organism="Salmonella typhimurium LT2"
/strain="SGSC1412"
/db_xref="taxon:99287"
/clone="STMD1"
<1..980
/note="r1ic"
rRNA

```

This sequence was finished as follows unless otherwise noted:
all regions were doubly stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one ml3 subclone.

rRNA	/product="23S ribosomal RNA" 1072..1191 /note="rriC"	gene	/translation="MTALLRVISLVISVYVVIIPPCGALGRGA" 4934..6580 /gene="11vG_1"
tRNA	/product="5S ribosomal RNA" 1244..1323 /product="tRNA-Asp" 1329..1401 /product="tRNA-Trp" complement(1506..2354) /gene="STM1_99" complement(1506..2354) /note="83% identity with E. coli hypothetical protein (Y1FA) (SP:P22788) and 95% identity with amino acids 1-80 of E. coli possible regulatory protein (PSSR) (SP:P27826)"	CDS	/gene="11vG_1" /note="STM1_91" /codon_start=1 /transl_table=11 /product="92% identity with E. coli acetoacetate synthase II (11vG) (SP:P00892) ; contains similarity to Pfam family PF00205 (Thiamine pyrophosphate enzymes), score=952.6, E=1.5e-295, N=1" /protein_id="AAE33483.1" /db_xref="GI:6960293"
gene	/transl_table=11 /evidence=not experimental /protein_id="AAE33488.1" /db_xref="GI:6960296"	gene	/translation="MNGAOWVYVHARAGVYGVYGGALMPYVDALYDGGVEHLIC RHHQDAALICVARSCTGVCYATSGPGATNLTJGLADALLDSVPVATGQVSAP PICTDAPQPDVYIGSLACTKHSFVNOISLEIPIAAEFPPANAGRPQVVDIPKD IOLASSELFPATTYANATPPQADVQQAQMLDQAKRPMLYGCGYQMAQVPAIRK FIAVTPVYCTGLGDAVEADPTYLGLGMHGTKANFPAYOEGDILLVAGARFDR VTKLNTFAPNASTVIMHIDIPAEKKRLQAHVAQGLDLSLPALQOEPKIDAKROSC AEIKARHARDPGEETIYAPILKQLSERKPADSVVTVDGQOHOMQSAQMTYTRPE NETSSGLGTMGFCGLPAAVGAQVAPVDVTCISGDSFAMVDELGVTRKQILPKI VILDNORLQMVQWQOLFQERSETTLTNDNPFLMASAEVIGQIHTRKDQVEAL DTMLAEGPYLLHVSIDLEENWVLPVPGASNSEMLKLS" 6580..6840 /gene="11vM"
CDS	/gene="y1fE" 2461..2799 /gene="y1fF" 2461..2799 /note="STM1_96" /codon_start=1 /transl_table=11 /product="93% identity with E. coli acetoacetate synthase (11vM) (SP:P13048)" /protein_id="AAE33482.1" /db_xref="GI:6960292"	CDS	/gene="11vM" 6580..6840 /gene="11vM" /note="STM1_90" /codon_start=1 /transl_table=11 /product="93% identity with E. coli acetoacetate synthase (11vM) (SP:P13048)" /protein_id="AAE33482.1" /db_xref="GI:6960292"
gene	/transl_table=11 /product="98% identity with E. coli hypothetical protein (Y1FE) (SP:P27827)" /protein_id="AAE33485.1" /db_xref="GI:6960295"	gene	/translation="MAGVNVASARFNEETLERVLRVRRHRCFOYCSMNNEAATDAQNI NIELTVASPRSYDLFLSOLSKVDVAHVAVICQSAATSDQIRA" 6858..7787 /gene="11vE"
CDS	/complement(2824..4344) /gene="y1fB" /note="STM1_95" /codon_start=1 /transl_table=11 /product="85% identity to E. coli hypothetical protein (Y1FB) (SP:P22787) ; contains similarity to Pfam family PF01078 (Magnesium chelatease, subunit ChII), score=351.8, E=7.6e-102, N=1" /protein_id="AAE33484.1" /db_xref="GI:6960294"	CDS	/gene="11vE" 6858..7787 /note="STM1_89" /codon_start=1 /transl_table=11 /product="S. typhimurium branched-chain-amino-acid transaminase (11vE) (SP:P15168)" /protein_id="AAE33481.1" /db_xref="GI:6960291"
gene	/complement(2824..4344) /gene="y1fB" /note="STM1_95" /codon_start=1 /transl_table=11 /product="85% identity to E. coli hypothetical protein (Y1FB) (SP:P22787) ; contains similarity to Pfam family PF01078 (Magnesium chelatease, subunit ChII), score=351.8, E=7.6e-102, N=1" /protein_id="AAE33484.1" /db_xref="GI:6960294"	gene	/translation="WTRKADYIMFNGEWMEDAKVHVMASHLYGTSEFEGRCYD SHKGPVFRHRHMQLRDSAKLYRFVVSQSIDELMEACRDVTRKNNTLSATYIRPLV VGDVGMGVNPPPGYTTDVIIAAPMGAVYLGAEALDQGDIAVMSNRAAPRITPAK AGNYSLSLVGSEARHRYOEGTALDVNYSISEGAGDELVEVKGVGLTFPPFTSAL PGTADRAIKLAKELGIEVRQVLSRESLTLADEVMSGTAEITPVRSVDGIQVSGS RCGPVARIKQOAFGLPTGFEDKWMGLDPVNS" 7948..9798 /gene="11vD"
CDS	/transl_table=11 /product="S. typhimurium d1hydroxy-acid dehydratase (11vD) (SP:P40810) ; contains similarity to Pfam family PF00920 (Dehydratase family), score=1023.9, E=3.5e-304, N=1" /protein_id="AAE33480.1" /db_xref="GI:6960290"	CDS	/gene="11vD" 7948..9798 /note="11vD" /note="STM1_88" /codon_start=1 /transl_table=11 /product="S typhimurium d1hydroxy-acid dehydratase (11vD) (SP:P40810) ; contains similarity to Pfam family PF00920 (Dehydratase family), score=1023.9, E=3.5e-304, N=1" /protein_id="AAE33480.1" /db_xref="GI:6960290"
gene	/transl_table=11 /product="100% identity to E. coli 11vGMDA operon leader peptide (11vL) (SP:P03060)" /protein_id="AAE33486.1" /db_xref="GI:6960296"	gene	/translation="MPKYSATTTGRCNNMAGARLARATQMSDFGKPIIAVANSFT QFVPGVHLRADGKLVIAEIDASGVAKENPTIANDGIGAMGGMATLSLRSLIND SVBYVNAHCADAVWCISNCKITPGMLASLINTPIYTVFSGGMEAGKTKLSDKIT KIDLVAMOGADPKVSDOSNOYERACTGSCGCMFTANSMCLTEALGDISOPN GSLIALHAPROKDELINACKRIVELTKRYEONDESLAPNLSAKAFENATGDIAG GSMYVLIHLLAAQEAETDPMDSIDKLSRKYPOLCKDAPSPQKRYKMEVDVRAAGVIG IIEEDLRAGLLRNRYKAVYIGLTLRPTLEQYDITVTDGDAVKKMFAPAGAGIRITQANS

QDCRMSLDDDRAGCIRSLLEVAYSKDGLAVLYGNFANGCIVKTAGVDSILKFTG
PAAYVESODDVEALLGKGVVEGVVIVRYRGSGQMEMLYPTSLKSMGLGKAC
ALTIDGRFSGGTSLGIHVSPEAGSTGTLIDSDGTALIDIPNKSIOLOLSEAEIA
ARREAGKRGDKAMTPKRNROVSPALRAVAYSLATSDKGVADVDSKSLIG"

gene 9801..11345

alignment_scores:
Quality: 845.00 Length: 207
Ratio: 4.424 Gaps: 2
Percent Similarity: 92.271 Percent Identity: 81.159

alignment_block:
US-09-466-935-2 x STYSTMD1/rev ...

Align seg 1/1 to reverse of: STYSTMD1 from: 1 to: 96086

```
1 MetThrLeuGluTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuTh 17
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61708 ATGACCTTGAATGAGTGCTGCGCTTCTACTGACCTCACTGCTGAG 61659

17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrSerLeuAsnH 34
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61658 TCTTCTCGGGCTCAGCGCCATCATCATGACGACGCTATCATACC 61609

34 IscLtyrProAla..GlyGlyValItyrCysTrpAlaSerAspArgThrG 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61608 ATGGAATTCGTGGCGGAGTGCCTTCTATCGCGGACTCGACCGGAGCG 61559

50 LysPserTyrCysAlaGlyTrpArg..GlyValGlyThrLeuPheSerAr 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61558 GGGATCATATCATCTGCTGAGCGCTGCGGCTGACGCTTCTTTCGCG 61509

66 gSerValIleAlaPheGluValLeuItyrTrpAlaGlyAlaIleAspLeuI 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61508 CTCGCTCTCGCTTCTTGAATTCGAAATGGGCTGCGCGCTTATCTTA 61459

83 IerPheGlyIleGlnIleTrpArgAlaAlaGlyAlaIleAspLeuIys 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61458 TCTGGCTGGGATTCGACGAATGGCGCGCGCGCTATGATCATCTCAT 61409

100 SerLeuAlaSerThrClnSerArgArgHisLeuPheGlnArgAlaValAlpH 116
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61408 ACCCTGCCCAACCCCAATCGCGGGCTGCTCAACGCGGATAT 61359

116 eValAsnLeuThrAsnProIysSerIleValPheLeuAlaIleuPheP 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61358 TGTATATCTACCAATCCCAAAAGTATGTCTTCTGCGCGCTGTTC 61309

133 rGlnIlePheIleMetProGlnIleProGlnMetGlnTyrIleValLeu 149
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61308 CGCAATTCATCATGCCGACGACCATGCGCGCATACCTCATCTTC 61259

150 GlyValThrThrIleValIleAspIleIleValMetIleGlyTyrAlaTh 166
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61258 GGCCTCACCAAGATGTGTGATATGCTGTGATGACCGGTACGCCAC 61209

166 rLeuAlaGlnArgIleAlaLeuTrpIleuIysGlyProIysGlnMetLysA 183
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61208 ACTGCGCAGCGCATTTGCCGCTGATTAAGACCAAGCAGATGAGG 61159

183 IalLeuAsnIlystIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeu 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61158 CGGTGAATTAAGCGTTTGTCTGTTTATATGCTGTAGGCGCTCTCTG 61109

200 AlAserAlaArgHisAla 205
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
61108 GCGTCGCGCAAGACACCGG 61091

seq_name: gb_h1g7.AC020870
```

seq_documentation_block:
LOCUS AC020870 296373 bp DNA HTG 17-FEB-2000
DEFINITION Mus musculus clone RP23-302J15, LOW-PASS SEQUENCE SAMPLING.

AC020870
AC020870.2 GI:6984372
HTG: HTGS_PHASB0.
KEYWORDS
SOURCE
ORIGINISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 256373)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 256373)
DOE Joint Genome Institute.
Direct Submession
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
on Feb 17, 2000 this sequence version replaced gi:6686438.
* NOTE: This record contains 198 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced
* however, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
753 1000: contig of 752 bp in length
gap of unknown length
753 1000: contig of 248 bp in length
gap of unknown length
1001 1604: contig of 604 bp in length
gap of unknown length
1605 2302: contig of 698 bp in length
gap of unknown length
2303 2699: contig of 397 bp in length
gap of unknown length
2700 3686: contig of 987 bp in length
gap of unknown length
3687 4099: contig of 413 bp in length
gap of unknown length
4100 4778: contig of 679 bp in length
gap of unknown length
4779 5571: contig of 793 bp in length
gap of unknown length
5572 5967: contig of 386 bp in length
gap of unknown length
5968 6763: contig of 796 bp in length
gap of unknown length
6764 7639: contig of 876 bp in length
gap of unknown length
7640 8442: contig of 803 bp in length
gap of unknown length
8443 9414: contig of 972 bp in length
gap of unknown length
9415 10236: contig of 822 bp in length
gap of unknown length
10237 10844: contig of 608 bp in length
gap of unknown length
10845 11504: contig of 660 bp in length
gap of unknown length
11505 12321: contig of 817 bp in length
gap of unknown length
12322 12652: contig of 331 bp in length
gap of unknown length
12653 14116: contig of 1464 bp in length
gap of unknown length
14117 15298: contig of 1182 bp in length
gap of unknown length
15299 16234: contig of 936 bp in length
gap of unknown length
16235 16376: contig of 142 bp in length
gap of unknown length

```
*      16377      16664: contig of 288 bp in length      gap of unknown length
*      16665      17631: contig of 967 bp in length      gap of unknown length
*      17632      19241: contig of 1610 bp in length      gap of unknown length
*      19242      20453: contig of 1212 bp in length      gap of unknown length
*      20454      20559: contig of 106 bp in length      gap of unknown length
*      20560      21117: contig of 536 bp in length      gap of unknown length
*      21118      21655: contig of 538 bp in length      gap of unknown length
*      21656      21818: contig of 163 bp in length      gap of unknown length
*      21819      22679: contig of 861 bp in length      gap of unknown length
*      22680      23579: contig of 900 bp in length      gap of unknown length
*      23580      24383: contig of 804 bp in length      gap of unknown length
*      24384      25612: contig of 1229 bp in length      gap of unknown length
*      25613      25957: contig of 345 bp in length      gap of unknown length
*      25958      27260: contig of 1303 bp in length      gap of unknown length
*      27261      28596: contig of 1336 bp in length      gap of unknown length
*      28597      29473: contig of 877 bp in length      gap of unknown length
*      29474      30372: contig of 899 bp in length      gap of unknown length
*      30373      31202: contig of 830 bp in length      gap of unknown length
*      31203      32688: contig of 1486 bp in length      gap of unknown length
*      32689      33431: contig of 743 bp in length      gap of unknown length
*      33432      33693: contig of 262 bp in length      gap of unknown length
*      33694      34616: contig of 923 bp in length      gap of unknown length
*      34617      35884: contig of 1268 bp in length      gap of unknown length
*      35885      36459: contig of 575 bp in length      gap of unknown length
*      36460      36996: contig of 537 bp in length      gap of unknown length
*      36997      37941: contig of 945 bp in length      gap of unknown length
*      37942      38682: contig of 741 bp in length      gap of unknown length
*      38683      39549: contig of 867 bp in length      gap of unknown length
*      39550      40792: contig of 1243 bp in length      gap of unknown length
*      40793      42585: contig of 1793 bp in length      gap of unknown length
*      42586      44115: contig of 1530 bp in length      gap of unknown length
*      44116      45300: contig of 1185 bp in length      gap of unknown length
*      45301      46417: contig of 1117 bp in length      gap of unknown length
*      46418      47815: contig of 1398 bp in length      gap of unknown length
*      47816      49310: contig of 1495 bp in length      gap of unknown length
*      49311      50304: contig of 994 bp in length      gap of unknown length
```

```
*      50305      50848: contig of 544 bp in length      gap of unknown length
*      50849      51828: contig of 980 bp in length      gap of unknown length
*      51829      52699: contig of 871 bp in length      gap of unknown length
*      52700      53365: contig of 666 bp in length      gap of unknown length
*      53366      54633: contig of 1268 bp in length      gap of unknown length
*      54634      55470: contig of 837 bp in length      gap of unknown length
*      55471      56358: contig of 888 bp in length      gap of unknown length
*      56359      56934: contig of 576 bp in length      gap of unknown length
*      56935      57672: contig of 738 bp in length      gap of unknown length
*      57673      58319: contig of 647 bp in length      gap of unknown length
*      58320      58529: contig of 210 bp in length      gap of unknown length
*      58530      59083: contig of 554 bp in length      gap of unknown length
*      59084      59906: contig of 823 bp in length      gap of unknown length
*      59907      60826: contig of 920 bp in length      gap of unknown length
*      60827      61084: contig of 258 bp in length      gap of unknown length
*      61085      61762: contig of 678 bp in length      gap of unknown length
*      61763      62638: contig of 876 bp in length      gap of unknown length
*      62639      62713: contig of 75 bp in length      gap of unknown length
*      62714      63593: contig of 880 bp in length      gap of unknown length
*      63594      64039: contig of 446 bp in length      gap of unknown length
*      64040      65076: contig of 1037 bp in length      gap of unknown length
*      65077      65366: contig of 290 bp in length      gap of unknown length
*      65367      65590: contig of 224 bp in length      gap of unknown length
*      65591      66041: contig of 451 bp in length      gap of unknown length
*      66042      67040: contig of 999 bp in length      gap of unknown length
*      67041      67560: contig of 520 bp in length      gap of unknown length

alignment_scores:
      Quality: 795.00      Length: 174
      Ratio: 4.789      Gaps: 3
Percent Similarity: 95.402      Percent Identity: 93.678

alignment_block:
US-09-466-935-2 x AC020870/rev ..
Align seg 1/1 to reverse of: AC020870 from: 1 to: 256373

1 MetPrnLaugIuTrPrpPhaLaTyLeuLeuThSerLiLeuTh 17
|||||
77559 ATGACCTTGAATGCTGCTTTCCTACCTGTCACATCATTTTAA 77510
|||||
17 rleu..SerProGlySerGlyAlaIleAspTherThSerLeuAs 33
|||||
77509 CCTGCTGCCCAAGGCTCTGTCATCAACACTATGACCACTGCTCAA 77460
|||||
33 nHtSLgTyTrProAla..GlyGlyValTyCyTrpAlaSerAspArgTh 49
|||||
```

77459 CCAGGTTATCGGCGGCGGTGCTATTGCTGGCTTCAGACCGGAC 77410
49 rGlyAspSerTyrCysAlaGlyTrpArgGlyVal GlyThrLeuPheSer 65
|||||
77409 TGGGATTCATTGTCGTGGTGGGCTGGGAGCGCTATTATTC 77360
66 ArgSerValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLe 82
|||||
77359 CGCTACATGATTCGTTGAGGTGAGTGGCAGCGCGCTACTT 77310
82 uIleTrpLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeu 99
|||||
77309 GATTGGCTGGGATTCGAGTGGCGCGCCCTGTGTCATATGACCTTA 77260
99 ySsSerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaVal 115
|||||
77259 AATCGCTGCGCTCTACTCATCCGTCGACATTTGTCACCGCGCAGTT 77210
116 PheValAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPh 132
|||||
77209 TTTGTGATCTCCACCAATCCCAAAAGTATTGTTCTTGGCGCGCTATT 77160
132 eProGlnPheIleMetProGlnGlnProGlnLeuMetClnTrpIleVal 149
|||||
77159 TCCGCAATTCATCATGCGCGCACGCGCAGTATGATATCGTGC 77110
149 euGlyValIThrThrIleValIleValIleValMetIleGlyTyrAla 165
|||||
77109 TCGGCGTACCACATTTGTGTCGATATATTGTCATGATGATTCACGCC 77060
166 ThrLeuAlaGlnArgIle 171
|||
77059 ACCGTGCTCAACAGGAGTC 77042
seq_name: gb_hhg8:AC022157
seq_documentation_block: 4833 bp DNA HTG 26-JAN-2000
LOCUS AC022157 Homo sapiens chromosome 5 clone RPl-280K18, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC022157
VERSION AC022157.1 GI:6758618
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 4833)
DOE Joint Genome Institute.
REFERENCE
AUTHORS Unpublished
JOURNAL 2 (bases 1 to 4833)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
* NOTE: This record contains 8 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 836: contig of 836 bp in length

* * * gap of unknown length
* 837 1566: contig of 730 bp in length
* * * gap of unknown length
* 1567 2612: contig of 1046 bp in length
* * * gap of unknown length
* 2613 2721: contig of 109 bp in length
* * * gap of unknown length
* 2722 2829: contig of 108 bp in length
* * * gap of unknown length
* 2830 3052: contig of 223 bp in length
* * * gap of unknown length
* 3053 4256: contig of 1204 bp in length
* * * gap of unknown length
* 4257 4833: contig of 577 bp in length.
FEATURES
source
1. 4833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPl-280K18"
BASE COUNT 1218 a 1298 c 1250 g 1055 t 12 others
ORIGIN
alignment_scores:
Quality: 751.50 Length: 211
Ratio: 4.175 Gaps: 5
Percent Similarity: 85.308 Percent Identity: 77.251
alignment_block:
US-09-466-935-2 x AC022157/rev ..
Align seq 1/1 to reverse of: AC022157 from: 1 to: 4833
10 LeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIleAs 26
:|||||
878 ATTGTGCTGCGCATTCCTCTATCTCAGCTCCGTGTC 840
26 nThrMetThrThrSerLeuAsnHisGlyTyrProAla..GlyGlyValTyr 42
|||||
839 AGGATGACCACTCGCTCAGACCGAGTATGCGCGCGGAGCGCTTA 791
42 rCyStrpAlaSerAspArgThrGlyAspSerTyrCysAlaGlyTrpArg. 58
|||||
790 TTGCTGGGCTTCAGACCGGAGTGGCGATTCATATTGCTGTTGGCGTG 741
59 GlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValLeuLys 75
|||||
740 GCGTTGGGATCGCTATTTCGCCGTCAGATGGCTTGAAGTGTGAA 691
75 strpAlaGlyAlaAlaTyrLeuIleTrpLeuGlyIleGlnGlnTrpArg 92
|||||
690 GTGGCGAGGCGCGCTTACTGATTTGGCTGGGATTCAGAGTGGCGG 641
92 IaAlaGlyAlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgArg 108
|||||
640 CCGTGTGTCGATTCGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 591
109 HisLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLysSerI 125
|||||
590 CATTTGTTCCAGCGCGCAGTTTGTGATTCACCAATCCCAAAAGTAT 541
125 eValPheLeuAlaIleLeuPheProGlnPheIleMetProGlnGlnProG 142
|||||
540 TGCTTTCTGGCGCGCTATTTCGCAATTCATGCGGACGACGCGCG 491
142 InLeuMetGlnTyr..... 146
490 AACGATGACAGTATACGTGCGCGCGCGCGGTCGAGCGGCGGTAGT 441
147 ...IleValLeuGlyValThrThrIleValAlaValIleIleValMetI 162
|||||
440 GTGCTGTGTTGCGAGTATGATGATTTTC.GTCGAGTGAATGCCGTTAA.. 394

162 eGIYrAlaThrlEuAlaGlnArgIleAlaLeuTrpIleYsGlyProL 179
|||||.....
393CCGGAGGCCCTTGCTCAACGATGTGCTATGATGAAGACCA 348
179 ysgInMetLysAlaLeuAnLysIlePheGlySerLeuPheMetLeuVal 195
|||||.....
347 AGCAGATGAAGCGCGCTGAATAGATTTCGCGCTGTTATGCTGTG 298
196 GLyAlaLeuLeuAlaSerAlaArgHisAla 205
|||||.....
297 GGAGCGCTGTACATCGCGCAGGCATGCC 268

seq_name: gb_hcg7:AC020970

seq_documentation_block:
LOCUS AC020970 212936 bp DNA HTG 10-FEB-2000
DEFINITION Mus musculus clone Rp23-252w21, WORKING DRAFT SEQUENCE, 144
unordered pieces.
ACCESSION AC020970
AC020970.1 GI:6691260
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 212936)
DOE Joint Genome Institute.
Unpublished
2 (bases 1 to 212936)
DOE Joint Genome Institute.
Direct Submission
Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation

NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
* 11047 12223: contig of 1177 bp in length
* gap of unknown length
* 12224 13492: contig of 1269 bp in length
* gap of unknown length

* 13493 14659: contig of 1177 bp in length
* gap of unknown length
* 14670 15886: contig of 1217 bp in length
* gap of unknown length
* 15887 17089: contig of 1203 bp in length
* gap of unknown length
* 17090 18308: contig of 1219 bp in length
* gap of unknown length
* 18309 19518: contig of 1210 bp in length
* gap of unknown length
* 19519 20710: contig of 1192 bp in length
* gap of unknown length
* 20711 21932: contig of 1222 bp in length
* gap of unknown length
* 21933 23606: contig of 1674 bp in length
* gap of unknown length
* 23607 25328: contig of 1722 bp in length
* gap of unknown length
* 25329 26649: contig of 1321 bp in length
* gap of unknown length
* 26650 27849: contig of 1200 bp in length
* gap of unknown length
* 27850 29135: contig of 1286 bp in length
* gap of unknown length
* 29136 30700: contig of 1565 bp in length
* gap of unknown length
* 30701 31877: contig of 1177 bp in length
* gap of unknown length
* 31878 33065: contig of 1188 bp in length
* gap of unknown length
* 33066 34267: contig of 1202 bp in length
* gap of unknown length
* 34268 35474: contig of 1207 bp in length
* gap of unknown length
* 35475 36650: contig of 1176 bp in length
* gap of unknown length
* 36651 37859: contig of 1209 bp in length
* gap of unknown length
* 37860 39072: contig of 1213 bp in length
* gap of unknown length
* 39073 40406: contig of 1334 bp in length
* gap of unknown length
* 40407 41630: contig of 1224 bp in length
* gap of unknown length
* 41631 42864: contig of 1234 bp in length
* gap of unknown length
* 42865 44200: contig of 1336 bp in length
* gap of unknown length
* 44201 46087: contig of 1887 bp in length
* gap of unknown length
* 46088 48227: contig of 2140 bp in length
* gap of unknown length
* 48228 49518: contig of 1291 bp in length
* gap of unknown length
* 49519 50910: contig of 1392 bp in length
* gap of unknown length
* 50911 52420: contig of 1510 bp in length
* gap of unknown length
* 52421 53613: contig of 1193 bp in length
* gap of unknown length
* 53614 55004: contig of 1391 bp in length
* gap of unknown length
* 55005 56300: contig of 1296 bp in length
* gap of unknown length
* 56301 57505: contig of 1205 bp in length
* gap of unknown length
* 57506 58686: contig of 1181 bp in length
* gap of unknown length
* 58687 60407: contig of 1721 bp in length
* gap of unknown length
* 60408 61632: contig of 1225 bp in length
* gap of unknown length
* 61633 62880: contig of 1248 bp in length

*			gap of unknown length
*	62881	64220:	contig of 1340 bp in length
*			gap of unknown length
*	64221	65420:	contig of 1200 bp in length
*			gap of unknown length
*	65421	66605:	contig of 1185 bp in length
*			gap of unknown length
*	66606	68682:	contig of 2077 bp in length
*			gap of unknown length
*	68683	70712:	contig of 2030 bp in length
*			gap of unknown length
*	70713	71895:	contig of 1183 bp in length
*			gap of unknown length
*	71896	73462:	contig of 1567 bp in length
*			gap of unknown length
*	73463	74806:	contig of 1344 bp in length
*			gap of unknown length
*	74807	76205:	contig of 1399 bp in length
*			gap of unknown length
*	76206	77580:	contig of 1375 bp in length
*			gap of unknown length
*	77581	79081:	contig of 1501 bp in length
*			gap of unknown length
*	79082	80314:	contig of 1233 bp in length
*			gap of unknown length
*	80315	81899:	contig of 1585 bp in length
*			gap of unknown length
*	81900	83140:	contig of 1241 bp in length
*			gap of unknown length
*	83141	84665:	contig of 1525 bp in length
*			gap of unknown length
*	84666	85888:	contig of 1223 bp in length
*			gap of unknown length
*	85889	87155:	contig of 1267 bp in length
*			gap of unknown length
*	87156	88435:	contig of 1280 bp in length
*			gap of unknown length
*	88436	89955:	contig of 1520 bp in length
*			gap of unknown length
*	89956	91277:	contig of 1322 bp in length
*			gap of unknown length
*	91278	92513:	contig of 1236 bp in length
*			gap of unknown length
*	92514	93917:	contig of 1404 bp in length
*			gap of unknown length
*	93918	95448:	contig of 1531 bp in length
*			gap of unknown length
*	95449	96800:	contig of 1352 bp in length
*			gap of unknown length
*	96801	97978:	contig of 1178 bp in length
*			gap of unknown length
*	97979	99179:	contig of 1201 bp in length
*			gap of unknown length
*	99180	100437:	contig of 1258 bp in length
*			gap of unknown length
*	100438	101827:	contig of 1390 bp in length
*			gap of unknown length
*	101828	103117:	contig of 1290 bp in length
*			gap of unknown length
*	103118	104550:	contig of 1433 bp in length
*			gap of unknown length
*	104551	105854:	contig of 1304 bp in length
*			gap of unknown length
*	105855	107523:	contig of 1669 bp in length
*			gap of unknown length
*	107524	109208:	contig of 1685 bp in length
*			gap of unknown length
*	109209	110430:	contig of 1222 bp in length
*			gap of unknown length
*	110431	111662:	contig of 1232 bp in length

alignment_scores: Quality: 734.00 Length: 206

	Ratio: 3.745	Gaps: 1
	Percent Similarity: 95.146	Percent Identity: 93.204
	alignment_block:	
	US-09-466-935-2 x AC020970/rev	..
	Align seg 1/1 to reverse of: AC020970 from: 1 to: 212936	
1	MetThrLeuGluThrPrpPheAlaTyrLeuLeuThrSerIleLeuTh	17
175539	ATGACCTTGAATGAGTGGTTGGCTACCTGTCGA.TGCATCATTTTAAG	175491
17	rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsn	34
175490	CCCTGTGCCAGGCTCTGTGCATCAACACTAT.ACACCTCGCTCAACC	175442
34	isgLyfyrProAla..GlyGlyValTyrCysThrPalaSerAparqThg	50
175441	ACGGTTATGCGGCGCGGTGGGCTGTATGTGGCT.TCAGACCGGACTG	175393
50	1AspSerTyrCysAlaGlyTyrPargLyValGlyThrLeuPheSerArg	66
175392	GGCATTCATATGTGTGTGGTGGCGTGGGAGCGCTATTTCCGCC	175343
67	SerValIleAlaPheGluValLeuLysTyrPalaGlyAlaIleTyrLeuI	83
175342	TCAGTATGTCCTTTGAACTGTGAAGTGCGC.GCGCGGCTTACTTGAT	175294
83	eTrrPleuGlyIleGlnGlnTrrPargAlaIleGlyAlaIleAspLeuLys	100
175293	TTCGTGGGAATCCACAGTGGCGCGCTGTGCACATTTGACCTTAAT	175244
100	eTrrLeuAlaSerThrGlnSerArgGlnIleLeuPheGlnArgAlaValPhe	116
175243	CGGTGGCCCTCTCATTCATGCGCTGCACATTT.TTCCAGCGCCAGTTT	175195
117	ValAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPheTr	133
175194	GTGAATCTCACCAATCCAAAGTATTGTGTCTG.GCGCGCTATTGCC	175146
133	oGlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuG	150
175145	GCAATTCATGATGCCGACACGCCGAACTGTG.CAGTATATCGTGTG	175097
150	1yValThrThrIleValValaIleIleValMetIleGlyTyrAlaThr	166
175096	GGGTACCACTATTGTGGTCGATATTATTGT.ATGATCGTTACGCCAC	175048
167	LeuAlaGlnArgIleAlaLeuThrPileLysGlyProLysGlnMetLysAl	183
175047	CTTGGTCACACGATGCTCTATGATTAAGA.CCAAGCAGATGAAGC	174999
183	AlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuA	200
174998	GCTGAATAGATTTTGGCTGTGTATTATGTG.CTGGGAGCGCTGTAG	174950
200	LaSerAlaArgHisAla 205	
174949	CATCGCGCAGCGCATCGC 174933	
seq_name:	gb_htg7:AC020874	
seq_documentation_block:		
LOCUS	AC020874 265383 bp DNA	HTG 16-FEB-2000
DEFINITION	Mus musculus clone Rp23-333M14, LOW-PASS SEQUENCE	SAMPLING.
ACCESSION	AC020874	
VERSION	AC020874.2 GI:6980207	
KEYWORDS	HTG; HTGS_PHASE0.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE	1 (bases 1 to 265383)	

AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 265383)
AUTHORS DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 16, 2000 this sequence version replaced gi:6686434.
* NOTE: This record contains 244 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 869: contig of 869 bp in length
* *
* 870 1181: contig of 312 bp in length
* *
* 1182 2202: contig of 1021 bp in length
* *
* 2203 2960: contig of 758 bp in length
* *
* 2961 3646: contig of 686 bp in length
* *
* 3647 4520: contig of 874 bp in length
* *
* 4521 4813: contig of 293 bp in length
* *
* 4814 5434: contig of 621 bp in length
* *
* 5435 6000: contig of 566 bp in length
* *
* 6001 6992: contig of 992 bp in length
* *
* 6993 7785: contig of 793 bp in length
* *
* 7786 8367: contig of 582 bp in length
* *
* 8368 9114: contig of 747 bp in length
* *
* 9115 9729: contig of 615 bp in length
* *
* 9730 10313: contig of 584 bp in length
* *
* 10314 10712: contig of 399 bp in length
* *
* 10713 11404: contig of 692 bp in length
* *
* 11405 12261: contig of 857 bp in length
* *
* 12262 13173: contig of 912 bp in length
* *
* 13174 13934: contig of 761 bp in length
* *
* 13935 14703: contig of 769 bp in length
* *
* 14704 15380: contig of 677 bp in length
* *
* 15381 15654: contig of 274 bp in length
* *
* 15655 16714: contig of 1060 bp in length
* *
* 16715 17249: contig of 535 bp in length
* *
* 17250 17863: contig of 614 bp in length
* *
* 17864 18508: contig of 645 bp in length

* *
* 18509 19193: gap of unknown length
* *
* 19194 19737: gap of unknown length
* *
* 19738 20600: gap of unknown length
* *
* 20601 21239: gap of unknown length
* *
* 21240 21903: gap of unknown length
* *
* 21904 22705: gap of unknown length
* *
* 22706 22948: gap of unknown length
* *
* 22949 23682: gap of unknown length
* *
* 23683 24364: gap of unknown length
* *
* 24365 25288: gap of unknown length
* *
* 25289 25954: gap of unknown length
* *
* 25955 26094: gap of unknown length
* *
* 26095 27015: gap of unknown length
* *
* 27016 27740: gap of unknown length
* *
* 27741 28060: gap of unknown length
* *
* 28061 28343: gap of unknown length
* *
* 28344 29016: gap of unknown length
* *
* 29017 29371: gap of unknown length
* *
* 29372 29586: gap of unknown length
* *
* 29587 29691: gap of unknown length
* *
* 29692 30501: gap of unknown length
* *
* 30502 31132: gap of unknown length
* *
* 31133 31249: gap of unknown length
* *
* 31250 32467: gap of unknown length
* *
* 32468 32878: gap of unknown length
* *
* 32879 34069: gap of unknown length
* *
* 34070 34694: gap of unknown length
* *
* 34695 35933: gap of unknown length
* *
* 35934 36920: gap of unknown length
* *
* 36921 37173: gap of unknown length
* *
* 37174 37282: gap of unknown length
* *
* 37283 38196: gap of unknown length
* *
* 38197 38452: gap of unknown length
* *
* 38453 38678: gap of unknown length
* *
* 38679 38835: gap of unknown length
* *
* 38836 39561: gap of unknown length

```

*      39562      40452: contig of 891 bp in length
*      40453      41223: contig of 771 bp in length
*      41224      41864: contig of 641 bp in length
*      41865      42690: contig of 826 bp in length
*      42691      42988: contig of 298 bp in length
*      42989      43612: contig of 624 bp in length
*      43613      44265: contig of 653 bp in length
*      44266      45263: contig of 998 bp in length
*      45264      46592: contig of 1329 bp in length
*      46593      47054: contig of 462 bp in length
*      47055      47193: contig of 139 bp in length
*      47194      48027: contig of 834 bp in length
*      48028      48670: contig of 643 bp in length
*      48671      49511: contig of 841 bp in length
*      49512      49895: contig of 384 bp in length
*      49896      50124: contig of 229 bp in length
*      50125      50776: contig of 652 bp in length
*      50777      50905: contig of 129 bp in length
*      50906      51260: contig of 355 bp in length
*      51261      52029: contig of 769 bp in length
*      52030      52294: contig of 265 bp in length
*      52295      53140: contig of 846 bp in length
*      *          *          gap of unknown length

```

alignment_scores:
 Quality: 637.00 Length: 235
 Ratio: 3.725 Gaps: 6
 Percent Similarity: 72.766 Percent Identity: 60.426

alignment_block:
 US-09-466-935-2 x AC020874 ..

Align seg 1/1 to: AC020874 from: 1 to: 265383

```

1 MetrrleuclutrrpphealaTyrrleuLeuThSerIleleuTh 17
55485 ATGACCTTAGAATGGTGGTTGCTACCTGACATCATCATATTAAAG 55534
17 rleuSerProGlySerGlyAlaIleasnPhrMetTrrrrSerLeuAsnH 34
55535 CTTTCGCCAGGCTCTGGTGCATCACTATGACCACCTCGCTCAACC 55584
34 IsGlyTyProAla..GlyGlyValTyrcyStrrrAlaSerAspArgThg 50
55585 ACGGTATTCGCGGCGGCGGCTATTCGCTTCAGACCGGACTG 55634
50 lyaasSerTyrcysAlaGlyTrpArgGlyAla GlyThrleuPheserAr 66
55635 GCGATTATATATGGCTGGTGGCGTGGGGTGGGACCGCTATTTCGCG 55684
66 gSeValIlealaphneGluValleuysTrpAlaGlyAlaAlaTyrleni 83

```

```

55685 CTCAGTATTCGGTTTGAAGTGTGAAGTGGCAGCGCGGCTTACTGA 55734
83 leTrleuclutrrpphealaTyrrleuLeuThSerIleleuTh 99
55735 TTTGGCTGGGAATCCAGTCAGTGGCGCGGCTGGTGCATTAATGCTTAA 55784
100 SerLeuAlaSerThrcGlnSerArgrHsLeuPheGlnArGAlaValAlp 116
55785 TCGCTGGCCTTACATCAATCGCGTGCACATTGTTCAGCGCGGAGTTT 55834
116 eValasleuThrAsnProLysSerIleValPheleuAlaIleu.... 131
55835 CGTGATTCACCAATCCAAAAGTATCGTGTCTGGCGGCGTACGTC 55884
131 ..... 131
55885 GGTACCGCAGGTGCTCGAATTTTGTCTGCTCGCTCGATAGTACG 55934
132 .....PheProGlnPheIle..... 136
55935 GCATTTGCGGAGAACAGCAGCGCAGCGCAGCGCAATACATTGGCT 55984
137 .....MetProGlnInProGlnLeuMetGlnTyrlay 148
55985 GGGCATCAGGTCGGTTCCTCCGGAAGCCAGGCTGACCTCAGTGGCG 56034
148 alLeuGlyValThrThrIleValAlaAspIleIleValMetIle.GlyTy 164
56035 CGAAGGATATTTACACCTGCGGCTCCGCGCATATATATTTCAGT 56084
164 rAlaTrleuAlaGlnArgIleAlaIleuTrIleLysGlyProLysGln 181
56085 AACTTCAGATCTTCTGCTTACTTACGCGCGGCAAGCATTACAT 56134
181 eLysAlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAla 197
56135 GCGGTACGCTTCATCTTATACGCGACGCTT...TTGACTGGCGCT 56181

```

seq_name: gp_htg11:AC026742

```

seq_documentation_block:
LOCUS AC026742 170178 bp DNA HTG 23-MAR-2000
DEFINITION Homo sapiens chromosome 5 clone CTD-3103E19, LOW-PASS SEQUENCE
SAMPLING.
ACCESSION AC026742.1 GI:7289992
VERSION AC026742.1
KEYWORDS HTG, HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 170178)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 5
Unpublished
2 (bases 1 to 170178)
DOE Joint Genome Institute.
Direct Submision
Submitted (23-MAR-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

```

* NOTE: This record contains 175 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone

```
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
*      1      635: contig of 635 bp in length
*      *      gap of unknown length
*      *      636      1342: contig of 707 bp in length
*      *      gap of unknown length
*      *      1343      1956: contig of 614 bp in length
*      *      gap of unknown length
*      *      1957      2656: contig of 700 bp in length
*      *      gap of unknown length
*      *      2657      2837: contig of 181 bp in length
*      *      gap of unknown length
*      *      2838      3485: contig of 648 bp in length
*      *      gap of unknown length
*      *      3486      4071: contig of 586 bp in length
*      *      gap of unknown length
*      *      4072      4571: contig of 500 bp in length
*      *      gap of unknown length
*      *      4572      4875: contig of 304 bp in length
*      *      gap of unknown length
*      *      4876      5560: contig of 685 bp in length
*      *      gap of unknown length
*      *      5561      6211: contig of 651 bp in length
*      *      gap of unknown length
*      *      6212      6915: contig of 704 bp in length
*      *      gap of unknown length
*      *      6916      7565: contig of 650 bp in length
*      *      gap of unknown length
*      *      7566      7879: contig of 314 bp in length
*      *      gap of unknown length
*      *      7880      8054: contig of 175 bp in length
*      *      gap of unknown length
*      *      8055      9082: contig of 1028 bp in length
*      *      gap of unknown length
*      *      9083      9199: contig of 117 bp in length
*      *      gap of unknown length
*      *      9200      10135: contig of 936 bp in length
*      *      gap of unknown length
*      *      10136      10450: contig of 315 bp in length
*      *      gap of unknown length
*      *      10451      11268: contig of 818 bp in length
*      *      gap of unknown length
*      *      11269      12291: contig of 1023 bp in length
*      *      gap of unknown length
*      *      12292      12948: contig of 657 bp in length
*      *      gap of unknown length
*      *      12949      13627: contig of 679 bp in length
*      *      gap of unknown length
*      *      13628      14182: contig of 555 bp in length
*      *      gap of unknown length
*      *      14183      15036: contig of 854 bp in length
*      *      gap of unknown length
*      *      15037      15215: contig of 179 bp in length
*      *      gap of unknown length
*      *      15216      15974: contig of 759 bp in length
*      *      gap of unknown length
*      *      15975      16596: contig of 622 bp in length
*      *      gap of unknown length
*      *      16597      16663: contig of 67 bp in length
*      *      gap of unknown length
*      *      16664      16954: contig of 291 bp in length
*      *      gap of unknown length
*      *      16955      17608: contig of 654 bp in length
*      *      gap of unknown length
*      *      17609      17762: contig of 154 bp in length
*      *      gap of unknown length
*      *      17763      17995: contig of 233 bp in length
*      *      gap of unknown length
*      *      17996      18075: contig of 80 bp in length
*      *      gap of unknown length
*      *      18076      18233: contig of 158 bp in length
*      *      gap of unknown length
*
*      18234      18936: contig of 703 bp in length
*      *      gap of unknown length
*      *      18937      19076: contig of 140 bp in length
*      *      gap of unknown length
*      *      19077      19299: contig of 223 bp in length
*      *      gap of unknown length
*      *      19300      19490: contig of 191 bp in length
*      *      gap of unknown length
*      *      19491      19705: contig of 215 bp in length
*      *      gap of unknown length
*      *      19706      19865: contig of 160 bp in length
*      *      gap of unknown length
*      *      19866      20017: contig of 152 bp in length
*      *      gap of unknown length
*      *      20018      20115: contig of 98 bp in length
*      *      gap of unknown length
*      *      20116      21069: contig of 954 bp in length
*      *      gap of unknown length
*      *      21070      22416: contig of 1347 bp in length
*      *      gap of unknown length
*      *      22417      23017: contig of 601 bp in length
*      *      gap of unknown length
*      *      23018      23192: contig of 175 bp in length
*      *      gap of unknown length
*      *      23193      23549: contig of 357 bp in length
*      *      gap of unknown length
*      *      23550      24450: contig of 901 bp in length
*      *      gap of unknown length
*      *      24451      24820: contig of 370 bp in length
*      *      gap of unknown length
*      *      24821      25794: contig of 974 bp in length
*      *      gap of unknown length
*      *      25795      26071: contig of 277 bp in length
*      *      gap of unknown length
*      *      26072      26717: contig of 646 bp in length
*      *      gap of unknown length
*      *      26718      26912: contig of 195 bp in length
*      *      gap of unknown length
*      *      26913      27578: contig of 666 bp in length
*      *      gap of unknown length
*      *      27579      28761: contig of 1183 bp in length
*      *      gap of unknown length
*      *      28762      28931: contig of 170 bp in length
*      *      gap of unknown length
*      *      28932      29143: contig of 212 bp in length
*      *      gap of unknown length
*      *      29144      30237: contig of 1094 bp in length
*      *      gap of unknown length
*      *      30238      30693: contig of 456 bp in length
*      *      gap of unknown length
*      *      30694      31274: contig of 581 bp in length
*      *      gap of unknown length
*      *      31275      31510: contig of 236 bp in length
*      *      gap of unknown length
*      *      31511      31680: contig of 170 bp in length
*      *      gap of unknown length
*      *      31681      31733: contig of 53 bp in length
*      *      gap of unknown length
*      *      31734      32506: contig of 773 bp in length
*      *      gap of unknown length
*      *      32507      33123: contig of 617 bp in length
*      *      gap of unknown length
*      *      33124      33855: contig of 732 bp in length
*      *      gap of unknown length
*      *      33856      34744: contig of 889 bp in length
*      *      gap of unknown length
*      *      34745      35489: contig of 745 bp in length
*      *      gap of unknown length
*      *      35490      35600: contig of 111 bp in length
*      *      gap of unknown length
*      *      35601      36330: contig of 730 bp in length
*      *      gap of unknown length
*      *      36331      36544: contig of 214 bp in length
```


gene
CDS

RAQRIRLEKLSRTQKRALRCRIGARRHLLFEDDAYNQNDLHLIRFPADSTIDHKRR
HHRP". .2722
/gene="VC0136"
2105. .2722
/gene="VC0136"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF93313.1"
/db_xref="GI:9654537"
/translation="MDIHVMTLVLTAVFSLAPSGTNSISGLSYGRHSIGAILI
GLQIGACHYIVGIGCALVAOASALPTILIKWIGAAYLYVMIGIQKRPRLPATIT
SHEISOALKKAVLINLTNPKSVTFVLALEPOFTDPDRHMFQVLGITVTYDIAI
VMFGTALAAGRGYIRSPNIMTRMKLFQSMFGCMGLATATA"

gene
CDS

2670. .3968
/gene="VC0137"
2670. .3968
/gene="VC0137"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF93314.1"
/db_xref="GI:9654538"
/translation="MVRCLMAEECCMLPRKKRPFEDDTWTYYARPIINAKRHILG
TELLRDESDAEPFYMDADRATRLVLENPLSLGTIPRLARSCTINFPHKSILRL
PLTPREDAIVEILETCPTDDLEAVQELSORGYLLADDEVYSPEWERFLPYVDLI
KIDIMAMLDKACEFNRGRLAQGSRRRFELAEVETDEDFQARKHAGTFQGGFFSKP
EIKORYYSPHVIAMOLFREVCSAPVDYVVERTLVADIALSYKLRFYNVMSDRIS
VSISRFAIYLIGODRKLRIFVSLAVASYSSKKRKELYNISLORAOFCOLMATHTHF
KAHPROATLIMFSYTDALIDPTSLIOLEYELOPLADVCKLAREREGPLGLLDELBCF
EKADMGVEGHCELGFDLEDYRQELLIEQRWSODINKRLI"
complement(3976. .4095)
/gene="VC0138"
complement(3976. .4095)
/gene="VC0138"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF93315.1"
/db_xref="GI:9654539"
/translation="MVCGPLSSYTLRWGLADHLSPNDLPLLPSTSLPROVKVH"
complement(4148. .4618)
/gene="VC0139"
complement(4148. .4618)
/gene="VC0139"
/note="similar to GB:I42023 SP:P45173 PID:1007363
PID:1221482 PID:1205585; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="DPS family protein"
/protein_id="AAF93316.1"
/db_xref="GI:9654540"
/translation="MATNLIGHDTQSOKLANLNLLANLYOVEYNTRGYHNIIQSK
EFPELHAFEELYDLOKIDELAIRLITLSPARNHSFGYKAAQIKENHDSIDGSS
SMOGVDSFSLHQORDILIELAGETGEGETSALMSDYIRQERTVLMNLAK"

gene
CDS

complement(4788. .5117)
/gene="VC0140"
complement(4788. .5117)
/gene="VC0140"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF93317.1"
/db_xref="GI:9654541"
/translation="MPFLLISLALLIITAAMVFSEHYRRHMOGENAPQRSAEVYTIID
KQAIDIPHLPDQDOEFMIYVVGKRLGPKRKEFOVGIIHYVALNPCKGVALTLOGQREFE
LHPALKR"

[illegible]

seq_name: gb_ba1:AE004937
seq_documentation_block: 10475 bp DNA BCT 30-AUG-2000
LOCUS AE004937 10475 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 498 of the complete genome.
ACCESSION AE004937 AE004091
VERSION AE004937.1 GI:9951553
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
1 (bases 1 to 10475)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S., Hutnagle,W.O., Kowalik,D.J., Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N., Polger,K.R., Kas,A., Lardig,K., Lam,R., Smith,K., Spencer,D., Mong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
JOURNAL MEDLINE 20437337
REFERENCE 2 (bases 1 to 10475)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,

115 alphaValAsnLeuThrAsnProlyserIleValPheLeuAlaIleLeu 131
1251 TCCTGCTTGTCATAGTGTGGCTGGCATTTGAATGGCGCTTGGTG 2300
49 ThrGlyAspSerTyrCysAlaGlyTTParg.GlyValGlyThrLeuphes 65
65 eraTsrSerValIleAlaPheGluValLeuYstrPalaGlyAlaIaTyr 81
2301 CACATATGGCTGGCTTCACCTCATCAATGATGATGGCGGCGCTTAT 2350
82 LeuIleTPrleuglyIleGlnGlnTPrPargAlaAlaGlyAlaIleasp 98
2351 TTAGTATGGCTGGGTATTCAGAAATGGCGATTTGAATGGCGCTTGGTG 2400
98 ulysSerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAla 115
2401 TACCACACTTCTCATGAACTCTCACAGCGCGCTGCTCGTAAAGCAG 2450
115 alphaValAsnLeuThrAsnProlyserIleValPheLeuAlaIleLeu 131
2451 TACTGATTAACCTGACCAATCCCAATCATCGTTTCTTACAGGCATTA 2500
132 PheProGlnPheIleMetProGlnInProGlnLeuMetGlnTyrIleVa 148
2501 TTTCGGCAGTTTATTTGATCCACATAGATCATTTGGCGCAGTTCTGGT 2550
148 lleGlyValThrThrIleValAlaIleIleValMetIleGlyTyr 165
2551 TCTGGGATACCAACCGTTGATGATGCGATTTGATGTCATGTCGGATCA 2600
165 IatTrLeuIaGlnArgIleAlaLeuTrpIleLysGlyPolySgIleMet 181
2601 CGGCACTGGCTGGCGACGACTAGTGCCTATATCTTACCTAACATATG 2650
182 LysAlaLeuAnLysIlePheGlySerLeuPheMetLeuValGlyAlaLe 198
2651 ACCAGAAATGAATAAAGTTTGGTTCGATGCTTATGGCTGCGGAATGCT 2700
198 uLeuAlaSerAlaArg 203
2701 GTTGGCTACCGCAA 2716
TITLE
JOURNAL MEDLINE 20437337
REFERENCE 2 (bases 1 to 10475)
Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P., Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,

FEATURES
source

1. 10475
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
153..743
/gene="PA5244"
153..743
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08629.1"
/db_xref="GI:9951554"
/translation="MMLOFLDQFGYFALFAGTFEGERTILVLAGEFLAERGWMQDVT
ILAFGLSYAGDQWYFLGRHGRRLARPRQKIGDALDVRHRHPLDMLVLSFRFV
YGRVWPAVIGSGYPPARYILNGAIVMAALGSAAYVYFVLEGMINKKYE
LWVLGGLIVIGLLIMLRKRTPRGNAGNADKSE"
908..1576
/gene="PA5245"
908..1576
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08630.1"
/db_xref="GI:9951555"
/translation="MHRKVAIVISGCGVYDGAIEHRSVITLRLSORGAEOCFAPNI
AOHHVYVHILTGEMPRESRNLVESARAREVDLEAREDEDALIVPGFCAKNL
SDRIRNGAQCVQPDVIALAKARAEAKRPGLGLICIRPAMAKIYKGAYVCTGNDQDT
ADALKQWAEHIDPVDIEVDSARKLVTPAYMALOSTAENASGINKRLDRLVELVD
QDKA"
1672..2145
/gene="PA5246"
1672..2145
/gene="PA5246"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08631.1"
/db_xref="GI:9951556"
/translation="MSETPSPALPDDYHOLVADFEQIPFNRMLGTRVASTIGRERYL
DLPEKDELIGNVOGILHGVYISLDDVTGAMALIGALRRHHEPEHEMARLSKIG
TIDRLVYLRGRGRTTAHAVPLRSKNKAVVYSELHSDGLVAVGSTYLCG"
complement(2130..2612)
/gene="PA5247"
complement(2130..2612)
/gene="PA5247"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08632.1"
/db_xref="GI:9951557"
/translation="MRLVIDADACPRAKELVCKFALKRKLEWMMVAGQOVYPPFAC
VRLVVSNGMAADYVLEADPECDLAICSDVPLADRLIKQVAAADLPKRFERDARKM
GDKIAMENLMLADLDDQGMGGCAGPYGERDROAFANALDRLTLRLQREADLRASQPHR
"

gene
CDS
complement(2679..4580)
/gene="PA5248"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

gene
CDS
/protein_id="AA008633.1"
/db_xref="GI:9951558"
/translation="MPPKILIMLPLVCAFSLGALADPAVDPSSQALHLLSYLAADYP
PVLAVAGKIVDPSEYOQVEFVGNLQALVLTLPMPREBELRGVASLROALEQRLPGR
DVALQARNLEKAVADIYQVVOPTPATPPPSRAAPVLAQOCAICGDAGKGGPGAGIGI
EPPNNLIDRQRLDHLISYDLRNVILGIVAGCDMDPAFDQDEORNDMLASYGNLSA
GSAODPKHAYPLATLATQTPAEVNEHGEAAASFRRLRAHPELQGRQOLIDYTNRA
ATLDPSPVPEGRDDQAVLSVAAYIEGFELVESSLDNDVADLRSTEKOLMAYROA
LADGIPETVOAQDELANGKLAENAKQGGSLSPSSISVSALLILIREGEALTVLA
ALLAFELRNTGQSSAVRGVAVGWGLFVAGFETWMLAAVYLDIGAQNELMEGFLSTPA
CMLVGLPWGHNVAIDRGAAAMQDYIRSSLVGGGGGFGFVAVLFEFSEYELFETFTST
IWMLOQPGHNAVIRIGGAATAVALLIGLAWVILRRSAPKPLDLFESINALLCALSYVE
AAGHYIALQENACVIGTRVPVPEFDWILKADAVLSAQAMALVAIALLVGRSIVTER
RRAANMAD"

gene
CDS
/protein_id="AA008634.1"
/db_xref="GI:9951559"
/translation="MLVSTWPAFLACWAISLSBAGAIASMSGCLQYGRAGYWNAL
GLQIALQAIIVAGVALLATSAFASLTKMPGVAVLYVLAQWQAPQALSTDG
BRPLQRLPLVLRGFLVNASNPKAVIFEMALVPOFIDPHOLAOVYLMGCTMIVDIL
IYMACYVTGLARVLVLRSPROOKLVNRTFASLVPGAAGLATVRRAPL"

gene
CDS
/protein_id="AA008635.1"
/db_xref="GI:9951560"
/translation="MEWLASPEIWAFLVLTALVLEIVLGIDNIIFISILVGRLPAPQAP
KARFELWAMGTRILLLISITWMAKRLTTDDLFLHFGGIGSRDLLEPFGGLPLIFST
MEIYSGMLAGEEOKSGGKGVFMGIIYOALIIDIFSLDSVITAEVYVGLVPMVNA
IIVISGVAMLAAGTISDFIEKHPSLKMLAISFLITVGTVLVAESEFVHVKGVYFMA
AFSLVVELNTRLRGALARKRGEDEPVLRKSGSP"

gene
CDS
/protein_id="AA008636.1"
/db_xref="GI:9951561"
/translation="MEDLOVLTQWSPPLVVAQOILILLVAMLLQRLVRRVTRIGE
RYVPAPPELLPLRGGLRWLIMWSALIVTLRGLVASVWLTALTGFVAVAAPFAM
SVLSMLFCALLITVGPFRIGRDELIDSDAKPCVTRGVAVAINLVTTLTEETLDDSP
GPLQIIPNSLSEFOKAVRRMRGPGFSTPKNGK"

gene
CDS
/protein_id="AA008637.1"
/db_xref="GI:9951562"
/translation="MIRLNLTLQGRFOQLLEAAEELTLHAGQKAGLIGANGAKSSLF
ALLRGQLQGDQGDCLPLPDMNRITAMRQVEJTLDRLLADVLDIDSRILREIOALAAE
AAHGSGLARHTEILDNADGYTADARAKRLLAGIGFSESEMERVDEFGSGWMRNL
AOALMCSDDLILDEPNTNLDLAILWLEMLGPGTETVLISHDDPFLDADVVDVHL
LENKRLITRYGSGYAFERTBAERLAAQOQAYEKQOARHMESTFIARAKATKARQA
OSRKALERLEPRLAPAVDSDPENSRESPKISRPLIDDEGRLGYDKAVLEKVKQ
LVYPARIGILAPNAGKSTLIKTLADGDELPGRLRGENTLAIQYRAQHQDLSDPQA
SPLLHDIQIAGEEEOATLKDFLGGDFRGVAVDEPVYINFGGGRARIALALAMQKPN
LTLLEPNTNLAAGEEEOATLMALOEFGAVLVVSHDRHLLKSTTDEFLIADGRVYF

[illegible]

[illegible][illegible]

*	35904	36602:	contig of 699 bp in length
*			gap of unknown length
*	36603	37110:	contig of 508 bp in length
*			gap of unknown length
*	37111	38059:	contig of 949 bp in length
*			gap of unknown length
*	38060	38730:	contig of 671 bp in length
*			gap of unknown length
*	38731	39790:	contig of 1060 bp in length
*			gap of unknown length
*	39791	40327:	contig of 537 bp in length
*			gap of unknown length
*	40328	40442:	contig of 115 bp in length
*			gap of unknown length
*	40443	41688:	contig of 1246 bp in length
*			gap of unknown length
*	41689	42688:	contig of 1000 bp in length
*			gap of unknown length
*	42689	43847:	contig of 1159 bp in length
*			gap of unknown length
*	43848	44511:	contig of 664 bp in length
*			gap of unknown length
*	44512	45780:	contig of 1269 bp in length
*			gap of unknown length
*	45781	47202:	contig of 1422 bp in length
*			gap of unknown length
*	47203	48647:	contig of 1445 bp in length
*			gap of unknown length
*	48648	49652:	contig of 1005 bp in length
*			gap of unknown length
*	49653	50485:	contig of 833 bp in length
*			gap of unknown length
*	50486	51527:	contig of 1042 bp in length
*			gap of unknown length
*	51528	53119:	contig of 1592 bp in length
*			gap of unknown length
*	53120	53623:	contig of 504 bp in length
*			gap of unknown length
*	53624	54669:	contig of 1046 bp in length
*			gap of unknown length
*	54670	55309:	contig of 640 bp in length
*			gap of unknown length
*	55310	56382:	contig of 1073 bp in length
*			gap of unknown length
*	56383	56625:	contig of 243 bp in length
*			gap of unknown length
*	56626	57345:	contig of 720 bp in length
*			gap of unknown length
*	57346	57505:	contig of 160 bp in length
*			gap of unknown length
*	57506	58908:	contig of 1403 bp in length
*			gap of unknown length
*	58909	59968:	contig of 1060 bp in length
*			gap of unknown length
*	59969	61377:	contig of 1409 bp in length
*			gap of unknown length
*	61378	62291:	contig of 914 bp in length
*			gap of unknown length
*	62292	63319:	contig of 1028 bp in length
*			gap of unknown length
*	63320	63818:	contig of 499 bp in length
*			gap of unknown length
*	63819	64823:	contig of 1005 bp in length
*			gap of unknown length
*	64824	65269:	contig of 446 bp in length
*			gap of unknown length
*	65270	65904:	contig of 635 bp in length
*			gap of unknown length
*	65905	66815:	contig of 911 bp in length
*			gap of unknown length

alignment_scores: 397.50 Length: 153
Quality: 397.50

Ratio: 3.312 Gaps: 6
Percent Similarity: 78.431 Percent Identity: 64.052
alignment_block:
US-09-466-935-2 x AC020885 ..

Align seg 1/1 to: AC020885 from: 1 to: 268294

```

55  AAGlyTPrgrgLyValGlyThrLeuPheSerArgSerValIleAlaPh 71
    |||||:|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
10824 GCCGGTTACCGG...TTTGGTGAACCGCACACATCGCATACGCTTACCAT 10870
71  eGIuValLeuLystrPaIaGlyAlaIaTyrrLeuIleTrpLeuGlyIleG 88
    :  ::  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
10871 CGGATCGCCAAA.....CATGCTCAGCAGTGTCTGCTCAATGTTA 10911
88  IngInTPrPaIaIaGlyAlaIleSpLeuLysserLeuAlaSerThr 104
    :|||  |||||:|||||  |||  |||  |||  |||  |||  |||  |||  |||
10912 TGCAGACGCGCGAGTAGGCGCA...GAAGTACCGACCGTAGGCGCCACT 10958
105  GlnSer...ArGArgHIsLeuPheGlnArgAla.ValIpheValAsnLeuT 120
    |||  |||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
10959 GGCAGTTGCGTACGGCAC.....CCGGCGGTGACGTGACTCTAAG 10999
120  hRAnPrOlysserIleValIlePheLeuAlaIaLeuPheProGlnPheIle 136
    :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|  :|
11000 TTCCCTCCCAAAAGATGATGTTCTGCGCGTATTCGCCAATTCATC 11049
137  MetProGlnInPrOgInLeuMetGlnTyrrIleValLeuGlyValIThrPh 153
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
11050 ATGCCGACACGCGCAACGTGATGATATTCGTGCTGCGCGCACAC 11099
153  rIleValIaAspIleIleValIleMetIleGlyTyrrAlaThrLeuAlaGlnA 170
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
11100 TATTGTGTCGATATTTATGTGATGATCGGTACGCCACCTTCTCAAC 11149
170  rGIleAlaLeuTrpIleLyGlyProLySgInMetLySaIaLeuAsnLyS 186
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
11150 GGATTTGCTTATGATTAAGACCAAGACGATGAGGCGCTGAAATTAAG 11199
187  IllePheLySerLeuPheMetLeuValGlyAlaLeuLeuAlaSerAlaAr 203
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
11200 ATTTTCGCTGTTGTTATGCTGTGTGGAGCGCTGTATGATCGGCGAG 11249
203  gHIsAla 205
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
11250 GCATGCG 11256

seq_name: gb_hlg7:AC020833

seq_documentation_block:
LOCUS      AC020833      62274 bp      DNA
DEFINITION Mus musculus clone RG-MBAC_40H2, LOW-PASS SEQUENCE SAMPLING.
ACCESSION  AC020833
VERSION    AC020833.1 GI:6686475
KEYWORDS   HTG; HTGS_PHASE0.
SOURCE     house mouse.
ORGANISM   Mus musculus.
REFERENCE  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS    Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
TITLE      1 (bases 1 to 62274)
JOURNAL    DOE Joint Genome Institute.
REFERENCE  Sequencing of Mouse
AUTHORS    Unpublished
TITLE      2 (bases 1 to 62274)
JOURNAL    DOE Joint Genome Institute.
AUTHORS    Direct Submission
COMMENT     Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely

```

* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 296: contig of 296 bp in length
* 297 1049: contig of 753 bp in length
* 1050 1812: contig of 763 bp in length
* 1813 2420: contig of 608 bp in length
* 2421 2953: contig of 533 bp in length
* 2954 3645: contig of 692 bp in length
* 3646 4766: contig of 1121 bp in length
* 4767 5296: contig of 530 bp in length
* 5297 5584: contig of 288 bp in length
* 5585 6454: contig of 870 bp in length
* 6455 6853: contig of 399 bp in length
* 6854 7566: contig of 713 bp in length
* 7567 8253: contig of 687 bp in length
* 8254 9058: contig of 805 bp in length
* 9059 9678: contig of 620 bp in length
* 9679 10272: contig of 594 bp in length
* 10273 10890: contig of 618 bp in length
* 10891 11695: contig of 805 bp in length
* 11696 12677: contig of 982 bp in length
* 12678 13130: contig of 453 bp in length
* 13131 13878: contig of 748 bp in length
* 13879 14026: contig of 148 bp in length
* 14027 14869: contig of 843 bp in length
* 14870 15738: contig of 869 bp in length
* 15739 16426: contig of 688 bp in length
* 16427 17161: contig of 735 bp in length
* 17162 17938: contig of 777 bp in length
* 17939 18715: contig of 777 bp in length
* 18716 19493: contig of 778 bp in length
* 19494 20212: contig of 719 bp in length
* 20213 20780: contig of 568 bp in length
* 20781 21499: contig of 719 bp in length
* 21500 21570: contig of 71 bp in length
* gap of unknown length

21571 22378: contig of 808 bp in length
22379 22940: contig of 562 bp in length
22941 23943: contig of 1003 bp in length
23944 24650: contig of 707 bp in length
24651 25686: contig of 1036 bp in length
25687 25820: contig of 134 bp in length
25821 26268: contig of 448 bp in length
26269 27152: contig of 884 bp in length
27153 28057: contig of 905 bp in length
28058 29002: contig of 945 bp in length
29003 29991: contig of 989 bp in length
29992 30780: contig of 789 bp in length
30781 31514: contig of 734 bp in length
31515 32152: contig of 638 bp in length
32153 32933: contig of 781 bp in length
32934 33042: contig of 109 bp in length
33043 34310: contig of 1268 bp in length
34311 34992: contig of 682 bp in length
34993 35821: contig of 829 bp in length
35822 37013: contig of 1192 bp in length
37014 38147: contig of 1134 bp in length
38148 38852: contig of 705 bp in length
38853 39840: contig of 988 bp in length
39841 40874: contig of 1034 bp in length
40875 42387: contig of 1513 bp in length
42388 43258: contig of 871 bp in length
43259 44365: contig of 1107 bp in length
44366 45674: contig of 1309 bp in length
45675 47108: contig of 1434 bp in length
47109 48115: contig of 1007 bp in length
48116 49871: contig of 1756 bp in length
49872 51661: contig of 1790 bp in length
51662 53728: contig of 2067 bp in length
53729 56784: contig of 3056 bp in length
56785 59220: contig of 2436 bp in length
59221 62274: contig of 3054 bp in length.
62274 location/Qualifiers
1. .62274

FEATURES
source

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RG-MBAC_40H2"

BASE COUNT      16965 a 13547 c 13251 g 18170 t      341 others
ORIGIN

alignment_scores:
    Quality: 270.50      Length: 131
    Ratio: 3.039      Gaps: 7
    Percent Similarity: 67.939      Percent Identity: 61.069

alignment_block:
US-09-466-935-2 x AC020833  ..

Align seg 1/1 to: AC020833 from: 1 to: 62274

82 LeuIeTrpLeuGlyIleGlnGlnTrpArgAlaIleGlyAlaIleAspLe 98
|||||
10213 TTGATATGCTTGCGC..... 10227

98 uLysSerLeuAlaSerThrGln_SerArgArgHisLeuPhe..... 111
::: ||||| |||::: |||||
10228 .....TGTTGCCCAACCCAAATCTCATCTTGACATTTCCCATGTGTT 10270

112 ...GlnArgAlaValPheValAsnLeuThrAsnProLysSerIleValPh 127
|||||
10271 GTTACGCGTCTGTG.....ATTCTCATCTCAAAAGATGTGTGTTTC 10311

127 eLeuAlaAlaLeuPheProGlnPheIleMetProGlnGlnProGlnLeuM 144
::: |||::: |||||
10312 TGGGGGGCGCTATT...TGCATATTCATCATGCCGCAACAGCGGCACTGA 10358

144 eGlnTrpIleValLeuGlyValThrThrIleValIleValAspIleIleVal 160
|||||
10359 TGCAGTATATGCTGCTGCGCTCGCCACTAT..GTGTGATATATTATGTG 10407

161 MetIleGlyTrpAlaThrLeuAla_GlnArgIleAlaLeuTrpIle...L 176
|||||
10408 ATGATCGGTGGCCACCCCTTGCTCANAAGATGCTCTATGCACTTAAN 10457

176 ySgLyProLySgLMetLysAlaLeuAsnLysIlePheGlySerLeuPhe 192
|||||
10458 AGGACCAACGACAGATGAAGCGCGCTGAATAGATTTCGGCTCGTGTGTT 10507

193 MetLeuValGlyAlaLeuLeuAlaSerAlaArgHisAla 205
|||||
10508 ATGCTGTGGGAGCGCTGTAGCATCGCGAGCATGCGC 10546

seq_name: gb_baz:AP003001

seq_documentation_block:
LOCUS      AP003001      345783 bp      DNA      ECT      03-FEB-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 8/21, complete
sequence.
ACCESSION  AP003001 BA0000012
VERSION    AP003001.1 GI:11994976
KEYWORDS   HTG.
SOURCE     Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM   Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.

REFERENCE
AUTHORS    Kaneko,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idegawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
TITLE      Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
JOURNAL    DNA Res. 7, 331-338 (2000)
REFERENCE  2 (bases 1 to 345783)
AUTHORS    Kaneko,T.
```

```
TITLE      Direct Submission
JOURNAL    Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp,
            URL:http://www.kazusa.or.jp/rhizobase/, Tel.81-438-52-3935,
            Fax:81-438-52-3934)
FEATURES
source     1..345783
            /organism="Mesorhizobium loti"
            /strain="MAFF303099"
            /db_xref="taxon:381"

BASE COUNT      64252 a 111029 c 106519 g      63983 t
ORIGIN

alignment_scores:
    Quality: 255.50      Length: 205
    Ratio: 1.907      Gaps: 4
    Percent Similarity: 65.366      Percent Identity: 33.171

alignment_block:
US-09-466-935-2 x AP003001  ..

Align seg 1/1 to: AP003001 from: 1 to: 345783

1 MetThrLeuGlnTrpTrpPheAlaTrpLeuLeuThrSerIleIleLeuTh 17
|||||
137079 ATGTGCTGGAGACTCTATGCCGCTTATGTTCTGCCGTCGATCGATCAT 137128

17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
|||||
137129 CCTGTGCCGGGACGACGACGACATGATCATGCCAAGCATCCGCC 137178

34 iSgLyTrpProAlaGlyIleGlyValTyrcyStrpAlaSerAspArgThgLy 50
|||||
137179 ACGGACCCGCGGCC...GGCCTCGCCAATGTCGCCGCGACGACGCGGG 137225

51 AspSerTyrcysAlaGlyTrpArgGlyValGly.....ThrLeuPheS 65
|||||
137226 CTTCGCATCATGATGCCATGTCGTCGCGCTCGCTCAACAGCTGATCGC 137275

65 rArgSerValIleAlaPheGluValLeuLysTrpAlaGlyAlaIleTyrl 82
::: |||||
137276 GGGCATGGGCGCACTGTTGTAATGGTCAGGCTGATGCGCGCCCTATTC 137325

82 euIleTrpLeuGlyIleGlnGlnTrpArgAlaIleGlyAlaIleAspLeu 98
|||||
137326 TGATCTGGATGGCGGTGCAGATGTTCCGCTCCAAGGCGACGCTGAATGCC 137375

99 LysSerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaVa 115
::: |||
137376 GACGGAAACCGCT...AGGAACCGCGCGCGCTTCTTCGACAGGCGCT 137422

115 lPheValAsnLeuThrAsnProLysSerIleValIlePheLeuAlaIleAsp 132
::: |||||
137423 GCTGTGGCGCTGAGACAATCCCAAGACCTGTGTTCTTCGCGCGCTTCT 137472

132 heProGlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrlIleVal 148
|||||
137473 TCCCGCAATTCATGCGCGGAGGCAACATGTTGCTGCTCAAAATGCTGCTC 137522

149 LeuGlyValThrThrIleValValAspIleIleValMetIleGlyTyrl 165
::: |||||
137523 ATGGGCTCGACCGCATGATCTTGCCCGCCATGTCGATGTGACCTACGC 137572

165 aThrLeuAlaGlnArgIleAlaLeuTrpIleGlySerProLysGlnMetL 182
::: |||
137573 GCTGCGCGCGCGCGCGCGGACGCGCTGCTGTCGCGC...AGCCGCAATCA 137619

182 ySAlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeu 198
|||||
137620 AGCTCATGTCCAGGATCAAGCGGACACTTCTGTGTTGGTGGCGGCGCTGTG 137669
```

Mon`May 7 11:08:03 2001

199 LeuAlaSerAlaArg 203
|||||
:::|
137670 CTGGCGTTTTCGAGG 137684

us-09-466-935-2.rge

Page 25

OM of: US-09-466-935-2 to: N_Geneseq_0401.* out_format : pfs
Date: May 6, 2001 3:43 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+ p2n.model -DPV=xlp
-O=/cgm2_1/USPTO.spool/US09466935/runat_03052001.075710.14843/app_query.fasta.1.529
-DB=N_Geneseq_0401 -QEMT=fastap -SUFFIX=ing -GAPOP=12.000
-GAPEXT=4.000 -MIMATCH=0.100 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.ccl -LIST=45 -DOCALIGN=200 -THR_SCORE=pcr
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pis
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09466935.@CGN1_1_213 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY
-WAIT -THREADS=1

Search information block:

Query: US-09-466-935-2
Query Length: 205
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 172.640000

Score List:

Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:294405 +			1061.00	2192.23	3.6e-114	1200
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A48442 +			1061.00	2191.94	3.7e-114	1231
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52689 +			195.50	359.17	9.7e-14	639
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:R06767 +			184.00	353.80	2.5e-12	720
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:V08553 +			184.00	353.80	2.5e-12	720
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:C74615 +			144.00	286.62	5.0e-08	401
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52688 +			126.50	244.32	1.1e-05	672
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52690 +			126.50	244.32	1.1e-05	672
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A61501 +			123.50	227.99	9.3e-05	1668
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A61502 +			123.50	227.99	9.3e-05	1668
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A48443 +			113.00	232.44	5.2e-05	840
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52691 +			113.00	232.44	5.2e-05	840
/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:R96816 +			89.50	152.69	0.0004	2374
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:V27579 +			85.50	145.01	3.87	2289
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:X13048 +			85.50	145.01	3.87	2289
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A81730 +			83.50	132.27	19.84	6590
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A81730 +			83.50	132.27	19.84	6590
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:V52687 +			82.50	142.46	5.37	1635
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:V27578 +			82.50	142.46	5.37	1635
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:V44288 +			82.00	137.81	9.75	2265
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:V44289 +			82.00	137.81	9.75	2265
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:X13075 +			82.00	109.73	357.43	28690
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:X13075 +			82.00	109.73	357.43	28690
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:F44697 +			80.50	132.45	19.38	2769
/SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:V74373 +			80.50	132.45	19.38	2769
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:X19826 +			80.00	138.55	8.87	1551
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X19827 +			80.00	138.55	8.87	1551
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X12151 +			79.50	137.62	10.00	1437
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X12151 +			79.50	137.62	10.00	1437
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:R21233 +			79.50	131.52	21.85	2494
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:R21233 +			79.50	131.52	21.85	2494
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A35111 +			79.50	111.97	268.23	14607
/SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A35113 +			79.50	111.97	268.23	14607
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:V05287 +			79.00	98.49	1.5e+03	49377
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X56244 +			79.00	98.49	1.5e+03	49377
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:R12223 +			78.50	114.75	14.45	1542
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:R12223 +			78.50	114.75	14.45	1542
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:V27577 +			78.00	129.34	28.91	2288
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:V27577 +			78.00	129.34	28.91	2288
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:R80948 +			78.00	117.00	34.62	2598
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:R80948 +			78.00	117.00	34.62	2598
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:V52731 +			77.50	144.83	3.96	513
/SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:X98132 +			77.50	144.83	3.96	513

/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:Q43846 + 77.50 111.08 295.08 10704
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:Q43846 + 77.50 111.08 300.57 10844
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:Q43844 + 77.50 110.40 327.86 11529
/SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:Q43847 + 77.50 109.84 352.45 12132
/SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:V52660 + 77.00 138.27 9.19 844

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:Z94405

seq_documentation_block:

ID Z94405 standard; DNA; 1200 BP.

XX Z94405;

XX 18-JUL-2000 (first entry)

XX E. coli rhtB gene imparting homoserine resistance.

XX Homoserine resistance; rhtB gene; L-homoserine; L-alanine;

XX L-isoleucine; L-valine; L-threonine; ss.

XX Escherichia coli.

XX Key Location/Qualifiers

FT CDS 557..1171

FT /*tag= "a
/note= "a DNA sequence corresponding to the coding
region (minus the stop codon) is
specifically claimed in Claim 3"

XX EP994190-A2.

XX 19-Apr-2000.

XX 20-SEP-1999; 99EP-0118581.

XX 13-OCT-1998; 98RU-0118425.

XX (AJIN) AJINOMOTO CO INC.

XX Llyshts VA, Zakataeva NP, Aleoshin VV, Belareova AV;

XX Tokhmakova II;

XX WPI: 2000-273530/24.

XX DR P-PSDB: Y79298.

XX Novel RhtB protein, useful for generation of L-homoserine resistance in
Escherichia bacteria and large-scale production of e.g. L-homoserine
and L-alanine

XX Claim 3; Page 10-11; 14pp; English.

This is the DNA sequence the novel rhtB gene of Escherichia coli
K-12. The rhtB gene maps at 86 min on the E. coli chromosome. The
gene participates in resistance to homoserine. Amplification of
the gene results in an improvement of the amino acid productivity
of E. coli. The invention provides: the rhtB protein (see V79298);
DNA encoding the rhtB, especially nucleotides 557-1171 of the
present sequence; a bacterium, especially of the genus Escherichia,
in which L-homoserine resistance is enhanced by amplifying the
copy number or increasing the expression rate of the rhtB DNA, the
DNA being carried on a multicopy vector or on a transposon; and a
method for producing an amino acid by cultivating the bacterium in
a culture medium to produce and accumulate the amino acid in the
medium, from which it is recovered. The method is used for the
production of L-homoserine, L-alanine, L-isoleucine, L-valine or
L-threonine (all claimed).

XX Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 other;

alignment_scores:

Quality: 1061.00 Length: 205
Ratio: 5.176 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-466-935-2 x 294405

Align seg 1/1 to: 294405 from: 1 to: 1200

```

1 MethTrileuGIunrrprrPhehAaIaTyrlLeuauwThrSlelleleuTh 17
557 ATGACCTTAGAATGCTGGTTGGCTAACCGCTGGACATCGATCATATTTTAAC 606
17 rLeuSerProGIySerGIyAlaIleasnThrMetThrThrSerLeuAsnH 34
607 GCTGCGCGACGCGCTCTGCTGCATTCAAACACTMTGACCACTCGCTCAACC 656
34 IsglyTyrrProAlaGlyIyValTyrrCysTrpAlaSerAspArgThgIy 50
657 ACGGTAATCCGCGCGGCGGCTGCTATTTGGTGCGCTTCAGACCGCACTGGC 706
51 AspSerTyrrCysAlaGlyTyrrTPAArgIyValGIyThrLeuPheSerArgSe 67
707 GATTCATATTTGCTGCTGGTTGGCTGGGGTTGGAGCGCATTTTTCCGCGTC 756
67 rValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrrLeuIleT 84
757 AGTATATTCGGCTTTGAATGTGTGAAGGGGACAGCGCGGCTACTTGATTTT 806
84 rPleuGIyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeuLysSer 100
807 GGGTGGGAATCCACCAATGGCGGCGCGCTGGTGCAATTCCTTAATAATCG 856
101 LeuAlaSerThrGlnSerArgArgThIleuPheGlnArgAlaValAlaPheVa 117
857 CTGGCGCTCTACTCAATGGCGTGACATTTGTTCCAGCGCGACGATTTTGT 906
117 IAsnLeuThraSnProLysSerIleValaPheLeuAlaAlaLeuPheProg 134
907 GAATCTCCACCACTCCAAAGATTTGTGTCTTGGCGGGCGCTAATTTCCGC 956
134 IAsnIleMetProGlnGlnProGlnLeuMetGlnTyrrIleValleuGIy 150
957 AATTCATCATTCGCCGACACAGCCGCAACTGATGCAATATGCTGCTGGCG 1006
151 ValThrThrIleValaIAspIleIleValaMetIleGlyTyrrAlaThrLe 167
1007 GTCACCACTAATTTGGTCGATATTTATTTGATGATGATCGATTACGCCACCT 1056
167 uAlaGlnArgIleAlaLeuTrpIleLysGIyProLysGlnMetLysAlaL 184
1057 TGTCTAACGGGATTCCTTATGATTAAGAGACCAAGCAGATGAAGGGCGC 1106
184 euAsnLysIlePheGlySerIleuPheMetLeuValGIyAlaLeuLeuAla 200
1107 TGAATTAAGATTTTGGGCTGCTTTTATCTGCTGGTGGAGACCGCTGTTTACA 1156
201 SerAlaArgHisAla 205
1157 TCGCGGAGCGCATGGC 1171

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/Na2000.DAT.A48442
seq_documentation_block:
ID A48442 standard; DNA; 1231 BP.
XX
AC A48442;
XX
DT 08-SEP-2000 (first entry)
XX
DE E. coli L-homoserine resistance gene, rhtB.
XX
KW L-homoserine resistance; L-homoserine synthesis; rhtB;
KW L-threonine; L-valine; L-leucine; ds.
XX

```

OS	Escherichia coli.	Location/Qualifiers
XX		557..1174
FH	Key	/*tag= a
FT	CDS	/product= "RntB"
FT		

PN EP1013765-A1.

PD 28-JUN-2000.

PF 20-DEC-1999; 99EP-0125406.

PR 23-DEC-1998; 98RU-0123511.

PA (AJIN) AJINOMOTO KK.

PI Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;

DR WPI; 2000-414602/36.

XX

PT to enhanced Rhtc protein activity, used to produce L-threonine,

[illegible][illegible]

Escherichia coli. This sequence may
be present in one or more

CC resistance on *E. coli* bacteria which would be useful for producing
CC a high yield of L-homoserine. L-homoserine resistance means that the
CC bacteria will be able to grow on a minimal medium containing
CC L-homoserine at a concentration at which the corresponding wild-type
CC strain would not grow. Since the transformed bacteria can grow on the
CC minimal medium, it can synthesise L-homoserine, which accumulates. The
CC accumulated amino acids can then be removed from the culture medium.
CC The bacterium of the present invention may also be used to synthesise
CC L-threonine, L-valine and L-leucine at increased levels.

Sequence 1231 BP; 291 A; 319 C; 311 G; 310 T; 0 other;

alignment_scores:

Quality: 1061.00

Length: 205

Ratio: 5.176

Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-466-935-2 x A48442 . .

Align seg 1/1 to: A48442 from: 1 to: 1231

1 MethTrpLauGIuTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuTh 17
 11 MethTrpLauGIuTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuTh 17
 557 ATGACCTTAGAATGATGGTGTGGCTCCTACCTGTCACATCGATATTTCAC 606
 17 rLeuSerProGIuSerGIyAlaAlIeasrTrpMetTrpThrSerLeuAsnH 34
 607 GGTGTGGCCAGGCTCTGGTGTGCATTAACACATATGACCACTCGGTCAAC 656
 34 IsGIyTYrFrcAlaGIyAlGIyAlTYrCYrTrpAlaSerAspArgTrpGIy 50
 657 ACGGTTATCCGGCCGGTGGCGCTCATTTGGCTGTGGCTTCAGACGGACGGC 706
 51 AspSerTYrCYsAlaGIyTYrPArgGIyAlGIyAlThrLeuPheSerArgSe 67
 707 GATCATATATGTGGCTGTGGCGTGGGGTGGGAGCGTATTTTCCCGCTC 756
 67 rAlIleAlaPheGIuAlIleuLysTrpAlaGIyAlAlAlATyrLeuIleT 84
 757 AGTGATGTGGCTTGAAGTGTGAAGTGGCGCGCGCTACTACTGATTT 806

```
84 rpleuglylleinglntrpargalaaiaaglyalaalleaspleuylsSer 100
|||||
807 GGCTGGGAATCCAGAGTGGCGCGCTGGTGCATTTGACCTTAATTCG 856
101 leuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValPheVa 117
|||||
857 CTGGCTTACTCAATCGCGTGCACATTTGTTCCAGCGCGGAGTTTGT 906
117 lnsLeuThrAsnProLysSerIleValPheLeuAlaLeuPheProG 134
|||||
907 GAATCTCACCAATCCAAAAGATTGTGTTCGCGCGCTAATTCGCG 956
134 lnpHeileMePProGlnGlnProGlnLeuMetGlnTyrIleValIleuGly 150
|||||
957 AATTTCATCATGCCGACAGCGCGCAACTGATGATATATGCTCTGGCG 1006
151 ValThrThrIleValValAspIleIleValMetIleGlyTyrAlaThrLe 167
|||||
1007 GTCCACCAATTTGTGTCATATTATTTGTATGATGATCGGTTACGCCCT 1056
167 uAlaGlnArgIleAlaLeuThrPleLysGlyProLysGlnMetLysAlaL 184
|||||
1057 TGTCTACAGCATTTGCTCTATGATTAAGACCAAGCATGTAAGCGCG 1106
184 euAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuAla 200
|||||
1107 TGAATTAAGATTTTCGCTGCTGTTTATGCTGTGGAGCGCTGTAGCA 1156
201 SerAlaArgHisAla 205
|||||
1157 TCGCGAGCATGCG 1171

seq_name: /SIBS2/gcgdata/geneseq/geneseqn/NA2000.DAT.A52689
seq_documentation_block:
ID A52689 standard: DNA; 639 BP.
XX
AC A52689:
XX
DT 03-JAN-2001 (first entry)
XX
DE Escherichia coli yeast gene.
XX
KW E. coli; yeast gene; amino acid production; excretion protein gene;
amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..639
FT /*tag= a
FT /product= "Yeast"
XX
PN EP1016710-A2.
XX
PD 05-JUL-2000.
XX
PE 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI livshits VA, zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
Tokhmakova II.;
XX
DR WPI: 2000-414802/36.
XX
P-PSDB: B01787.
XX
PT Increased production of L-amino acids by an Escherichia bacterium
comprises increasing the expression amount of an L-amino acid excretion
protein -
```

```
XX
PS Disclosure; Page 19-20; 29pp; English.
XX
CC The present sequence is the yeast gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in lysine, alanine, valine, histidine,
CC isoleucine, glutamic acid and proline is achieved if multiple copies of
CC the gene are transfected into a bacterium. The bacterium used is E. coli.
XX
SQ Sequence 639 BP; 142 A; 128 C; 161 G; 208 T; 0 other;

alignment_scores:
Quality: 195.50 Length: 213
Ratio: 1.387 Gaps: 10
Percent Similarity: 66.197 Percent Identity: 28.638

alignment_block:
US-09-466-935-2 x A52689 ..

Align seg 1/1 to: A52689 from: 1 to: 639

3 leuGlnTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuThrLeuSe 19
|||||
22 CTGAATTAAGTGG...ACCTATCTGTTGGGCGCATTTTATGCTGTG 68
19 rProGlnSerGlyAlaIle.....AspThrMetThrThrSerL 32
|||||
69 GCCAGGCGCAAAATACCGCTGTTGACTCAAAATATGATGATGACGGTA 118
|||||
32 euAsnHisGlyTyrProAlaGly..GlyValTyrCysTrpAlaSerAsp 47
|||||
119 TGAATTAAGCTTATCTCTCGCGCTGGTGTATTT..... 153
48 ArgThrGlyAspSerTyrCys.....AlaGlyTrpArgGlyValGlyTh 62
|||||
154 ..ATTGGGAGTGGGATGATGATGTTCTGGCATGGGCTGAGCGGAC 200
62 lPhePheSerArgSerValIleAlaPheGlnValLeuLysTrpAlaGly 79
|||||
201 ATTAAATTAAGACACACCGCATATATCAACATGTCATGATCTGTGTC 250
79 lAlaTyrLeuIleTrpLeuGlyIleGlnGlnTrpArgAlaIleGlyAla 95
|||||
251 CGTTTATTTGCTCTATCTGCGGAGTAAATTTCTTACGCGACC..... 294
96 lIleAspLeuLys.....SerLeuAlaSerThrGlnSerArgArgHl 109
|||||
295 .....CTGAAGGGTAAATATGAGCGGCAATCCGATGAGCGCCATA 338
109 s.....LeuPheGlnArgAlaValPheValAsnLeuThrAsnProLys 124
|||||
339 CGGTGCTATTTTAAACGCGGCTTAATTTGAGCTCATGATACCGAAG 368
124 erIleValPheLeuAlaLeuPheProGlnPheIleMetProGlnGln 140
|||||
389 CCATTTGTCTATGTCGCTTTTTCGACAGTTATGCAATGAATGAGCC 438
141 ProGlnLeu...MetGlnTyrIleValLeuGlyValThrThrIleValVa 156
|||||
439 CCACATACGSGGAATTTCACTCTTATCTGCGGCGAGCGTGAACGCT 488
156 lAspIleIleValMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaL 173
|||||
489 GAGTTTCTGTATTTGAGCTCTCGATTAATATCTGCGCTTTGTCACGC 538
173 euTrpIleLysGlyProLysGlnMet...LysAlaLeuAsnLysIlePhe 188
|||||
539 AGTACATACGTACCAAAAGAAACCTGCTAAAGTTGCGACACTCAGTAT 568
189 GlySerLeuPheMetLeuValGlyAlaLeuLeuAlaSer 201
```

```

seq.name: /SIDS2/gcgdata/geneseq/geneseqn/A1196.DAT:T06767
seq-documentation_block:
ID T06767 standard; DNA: 720 BP.
XX
AC T06767;
XX
DT 01-JUL-1996 (first entry)
XX
DE Mel-linked mlga gene.
XX
KW Marine mela; selectable marker; oyster larva settlement;
KM pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
KW mlga gene; ss.
OS Shewanella colwelliana strain LSTDF.
XX
FH Key Location/Qualifiers
FT RBS 154..159
FT RBS /*tag= a
XX
PN US5474933-A.
XX
PD 12-DEC-1995.
XX
PE 21-MAR-1990; 90US-0496804.
XX
PR 08-NOV-1993; 93US-0148945.
PR 21-MAR-1990; 90US-0496804.
PR 10-NOV-1992; 92US-0374837.
XX
PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI Fuqua WC, Welner RM;
XX
DR WPI: 1996-039515/04.
DR P-PSDB; R87527.
XX
PT Novel gene encoding marine mela from Shewanella - useful as
PT selectable marker in genetic engineering and for inducing larval
PT oyster settlement
XX
PS Example 7; Fig 14; 47pp; English.
XX
CC The 5' end of an open reading frame (T06767) was identified
CC directly downstream of the Shewanella colwelliana mela gene
CC (T06766), and was designated mlga (mel-linked gene). The
CC role of the encoded protein (R87527) was unclear as deletion
CC subcloning in E. coli demonstrated that only mela was required
CC for melanogenesis.
XX
SQ Sequence 720 BP; 180 A; 165 C; 165 G; 210 T; 0 other;
XX

Alignment_scores:
Quality: 184..00 Length: 210
Ratio: 1.546 Gaps: 4
Percent Similarity: 56.667 Percent Identity: 27.619

alignment_block:
US-09-466-935-2 x T06767 ..

Align seg 1/1 to: T06767 from: 1 to: 720

9 TyrLeuThrSerIleLeuThrLeuSerProGlySerGlyAlaI1 25
:::| | | ::::::::::::::::::::| | | |
122 TTTTACGCAGCCTTTTTCGTCAATTACACCAGTAGTGATGAC 171
25 ekaSnThreThrThrSerLeuasnHISgLYTYrPrOAlagLyGLyValr 42
* ::::| | | | | :::::::::::::: | | |

```

[illegible]

```
PR 21-MAR-1990: 900S-0496804.
PR 10-NOV-1992: 920S-0974837.
PR 08-NOV-1993: 930S-0148945.
XX
PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI Fugua WC, Weiner RM.
XX
DR WPI: 1999-058995/05.
DR P-PSDB: W73358.
XX
PT Mela melanin protein from marine bacteria - useful as UV blocker in
PS e.g. cosmetics
XX
PS Example 7, Fig 13, 57pp: English.
XX
XX This sequence encodes the Shewanella colwelliana M1ga protein, which is
CC related to the Mela protein of the invention. The invention relates to
CC Mela proteins from marine bacterium of the genus Shewanella, Vibrio or
CC Hyphomonas. The Mela protein is active in catalysing production of
CC melanin synthesis, as well as components of marine exopolysaccharides.
CC The Mela protein can be used as a UV blocker in sunscreen, pigment and
CC dye compositions useful in cosmetic applications, as tints for glass or
CC in paints and coatings for plastics, synthetic resins and fabrics, rubber
CC and wood.
XX
SQ Sequence 720 BP, 180 A, 165 C, 165 G, 210 T, 0 other;

alignment_scores:
    Quality: 184.00      Length: 210
    Ratio: 1.546         Gaps: 4
    Percent Similarity: 56.667   Percent Identity: 27.619

alignment_block:
US-09-466-935-2 x V08533 ..

Align seg 1/1 to: V08533 from: 1 to: 720

9 TyrLeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIle 25
   ::||| ||| ::::::::::::::::::::|||
122 TTTTACCGACCTTTTTCGTCTCAATTACACGAGTATGTGTATGAC 171
   ::||| ||| ::::::::::::::::::::|||
25 eAsrThmEthrThrSerLeuasnHISGlyTyrProAlaGlyGlyValr 42
   ::||| ||| ::::::::::::::::::::|||
172 ACTGGCATGACTCTCGTATGAGTATCGGTG. 205
   ::||| ||| ::::::::::::::::::::|||
42 yrcYstYPAlaSerAspArgThrGlyAspSerTyrCysAlaGlyTyrParg 58
   ||||| ||| ::::::::::::::::::::|||
206 .....CGCCGAAC.....TTATGATG 223
   ||||| ||| ::::::::::::::::::::|||
59 GlyValGlyThrLeuPheSerArgSerValIleAla..... 70
   ||||| ||| ::::::::::::::::::::|||
224 ATGGTTGGTGTAGCTAGCAGGCGGTGCCCTCGGCGATTCGCCCGCTAAT 273
   ||||| ||| ::::::::::::::::::::|||
71 .....PheGluValLeu. 75
   ||||| ||| ::::::::::::::::::::|||
274 GGGTGTGCCAGTATGATGCTGAACATTCACAACCTCTCGATATTTTAA 323
   ||||| ||| ::::::::::::::::::::|||
75 ySTrPALaGlyAlaIleTyrLeuIleTrrPleuGlyIleGlnGlnTrrParg 91
   ||||| ||| ::::::::::::::::::::|||
324 AATGGGTGCGTGGGCTCTATCTTGTGTACATCGCATTTAGCATGTGGCG 373
   ||||| ||| ::::::::::::::::::::|||
92 AlaAlaGlyAlaIle...AspLeuLysSerLeuAlaSerThrGlnSerAr 107
   ||||| ||| ::::::::::::::::::::|||
374 GCCAAGAGCAAAATGGCCAACTTGACATACCTCCAGTCAGATCACTAA 423
   ||||| ||| ::::::::::::::::::::|||
107 gArGHisLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLys 124
   ||||| ||| ::::::::::::::::::::|||
424 TCGACCGCTAATTAACCTCAAGGCTTGTACCGCAATTGCTAATCCAAAG 473
   ||||| ||| ::::::::::::::::::::|||
124 eTlleValPheLeuAlaIleLeuPheProGlnPheIleMetProGlnGln 140
   ::||| ||| ::::::::::::::::::::|||
```

```
474 GGTGGGCTTTATGATGATCGCTGCTCCGCCCTTTATATCAGCGTTGACCA 523
141 ProGlnLeuMetGlnTyrIleValLeuGlyValThrThrIleValAla 157
   ||| ::||| ||| ::|||
524 GCGATTGGACCAACAATTATGTATGATCTGCAATTATATATGATGACGA 573
157 PHeIleValMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaLeu 174
   ::||| ||| ::||| ||| ::|||
574 GTTCTTCAGCATGCTTGTCTTATGCGAGCGG.CGGAACAACCTTAACGT 622
174 rPHeLysGlyProLysGlnMetLysAlaLeuAsnLysIlePheGlySer 190
   ::||| ||| ::||| ||| ::|||
623 TTTTAAGTCAGAGCGCATCAATCAATGATGATCAACCGCATGACAGGAGT 672
191 LeuPheMetLeuValGlyAlaLeuLeuAla 200
   ||||| ||| ::||| |||
673 TTAAATGATCTGTGTGCTTATGTGTGCGG 702

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.C74616
seq_documentation_block:
ID C74616 standard; cDNA; 401 BP.
XX
AC C74616;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF171 polynucleotide sequence SEQ ID NO:341.
XX
XX Human: open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnerrary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; catilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
OS Homo sapiens.
XX
PN W0200058473-A2.
XX
PD 05-OCT-2000.
XX
PE 31-MAR-2000; 2000MO-US08621.
XX
PF 31-MAR-1999; 99US-0127607.
XX PR 02-APR-1999; 99US-0127636.
XX PR 05-APR-1999; 99US-0127728.
XX PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR P-PSDB: B40407.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 5; Page 607-608; 5507pp: English.
XX
XX C74446 to C77606 encode the proteins given in B40237 to B43397, which
XX represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnerrary;
```

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORF-associated disorder. The nucleic acids can be used to express ORF
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

S0 Sequence 401 BP; 57 A; 121 C; 124 G; 98 T; 1 other;

alignment_scores:
Quality: 144.00 Length: 129
Ratio: 1.694 Gaps: 4
Percent Similarity: 65.891 Percent Identity: 31.008

alignment_block:
US-09-466-935-2 x C74616 ..

Align seg 1/1 to: C74616 from: 1 to: 401

```
79 ALAAlaIreuleuIletrIleuglyIleGlnGlnItrPArgAlaIaGlyA 95
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7  GCGGCCCTACCTGCTGTAACCTGGCCCTATGCCACTGGCGCTGACCGCTCGC 56
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
95 alleAspleuLysSerleuAlaSerThrGlnSerArgArgHisLeuPheG 112
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 CTTTGCAATGACGACACGCCGCGACAGTGGACCGCGCGCAGCTGATCC 106
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
112 lnrArgAlaValPheValAsnLeuThrAsnProLysSerIleValPheLeu 128
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
107 TGGGTGGCTCTTGCTGAACATCTTAACCCCAAGCTGACATTTTCTTC 156
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
129 ALAAlaLeuPheProGlnPheIleMetPro.....GlnGlnProGlnLe 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
157 CTGGCCCTCTCGCTCATTCCTTAACGCCAGCGGCGGACCGCGCCCTT 206
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 lmetGlnTrIleValIleuGlyValThrThrIleValAlaIleIleIle 160
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
207 G...CAGATGCTGTACTGAGCGCGCTGTTCATGCGCATGACGCTGCAG 253
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
160 almetIleGlyTrAlaThrIleAlaGln.....ArgIleAlaLeuTrp 174
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
254 TgTTTGGCTGTATGGCTGTGTGGCAATGTGTTGCTGTGCACAG... 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
175 lIleGlyGlyProLysGlnIleTysAlaLeuAsnLysIlePheGlySerle 191
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
301 GTGGAGTGGCGACAGTGTGCAGAACTGCTGCGACGAGCTTTTCCACGCG 350
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
191 uPheMetLeuValGlyAlaLeuLeuAlaSerAlaArg 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 CTTTGGCGGCGCTGGGCTTGAACCTTGCCTTTCGCGCAG 387
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52688

seq_documentation_block:

ID A52688 standard; DNA; 672 BP.

AC A52688;

XX 03-JAN-2001 (first entry)

XX Escherichia coli yahn gene.

```
XX
KW E. coli; yahn gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..672
FT /tag= a
FT /product= "yahn"
XX
PM EP1016710-A2.
XX
PD 05-JUL-2000.
XX
PF 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhnakova IL;
XX
DR WPI: 2000-414802/36.
DR P-PSDB; B01786.
XX
PT Increased production of L-amino acids by an Escherichia bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX
PS Disclosure; Page 17-18; 29pp; English.
XX
CC The present sequence is the yahn gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in lysine, glutamic acid and proline is
CC achieved if multiple copies of the gene are transfected into a bacterium.
CC The bacterium used is E. coli.
XX
S0 Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;
```

alignment_scores:
Quality: 126.50 Length: 201
Ratio: 1.037 Gaps: 6
Percent Similarity: 60.697 Percent Identity: 21.891

alignment_block:
US-09-466-935-2 x A52688 ..

Align seg 1/1 to: A52688 from: 1 to: 672

```
1 MethIreugIuTrpPheAlaTrIleuLeuThr.....SerI 14
   ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
34 ATCACTATGATGCTTTGCAATGCCGTTTACCTGACCGCTAGACCTGTCGT 83
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
14 eIleLeuThrIleuSerProGlySerGlyAlaIleAsnThrMetThrThrs 31
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 GATTACTTTTATTATCCGGAGGCCAATCTCTTGTGTGTGTGTGATCAACCA 133
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
31 erLeuAsnHISGlyTrProAlaGlyGlyValIlyrCysTrpAlaSerAsp 47
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 GCCGTGGCTTCGCGTGCAGCGCA...GGGCTGTGACCGGCGTGGCGGTG 180
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
48 ArgThrGlyAspSerTrpCysAlaGlyTrpArg.....GlyValGlyTh 62
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 GCGCTGGGCGATGATTTTATTCGGGCTTGGCTTGTGTGTGTGTGCAAC 230
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
62 rIleuPheSerArgSerValIleAlaPheGlyValLeuIlyTrpAlaGlyA 79
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```


196 GLYAlaLeuLeu 199
|||||
544 CTGGCCCTGTGG 555

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:A61501

seq_documentation_block:

ID A61501 standard; DNA: 1668 BP.

AC A61501;

XX 23-OCT-2000 (first entry)

DE A. vitis hypersensitive response elicitor-encoding DNA, SEQ ID NO:6.

KW Hypersensitive response elicitor protein; HR: disease resistance;
KW insecticide; fungicide; antiviral; bactericide; growth enhancer;
KW stress resistance; transgenic plant; ds.

XX Agrobacterium vitis.

PN W0200028056-A2.

XX 18-MAY-2000.

PE 05-NOV-1999; 99MO-US26079.

PR 06-NOV-1998; 980S-0107387.

PA (CORR) CORNELL RES FOUND INC.

P1 Burr TJ, Herlache TC, Zhang H;

XX WPI: 2000-376567/32.

DR P-PSDB; B11630, B11631, B11632, B11633, B11634.

PT New protein from Agrobacterium vitis, useful e.g. for imparting
PT resistance to disease or stress to plants, is involved in production of
PT a hypersensitive response -

PS Claim 7; Page 82; 157pp; English.

CC Sequences A61501-A61524 represent nucleotide sequences encoding
CC Agrobacterium vitis hypersensitivity response (HR) elicitor proteins
CC (B11630-B11688). The HR is a rapid, localised necrosis that is
CC associated with the active defence of plants against many pathogens, and
CC occurs when a pathogenic organism interacts with a nonhost plant (i.e.
CC one in which intracellular bacterial growth and disease development do
CC not occur). Like other HR elicitors, the A. vitis elicitor functions in
CC non-host plants by causing a rapid hypersensitive response that results
CC in walling-off and killing of the pathogen. On grape plants, the A. vitis
CC elicitor induces a restricted necrosis of tissues, resulting in the death
CC of plant cells and induction of pathogen resistance. A. vitis HR elicitor
CC proteins, in non-infectious form, are used to treat plants or their seeds
CC to impart resistance to disease, such as those caused by fungi, bacteria
CC or viruses; and to enhance growth, e.g., to increase yield or to provide
CC earlier germination or maturation. The proteins can also be used to
CC control insects, to impart resistance to environmental stresses, e.g.,
CC cold, and to improve nutritional value, e.g., altered oil content. The
CC same effects can be produced by producing transgenic plants or seeds by
CC incorporation of DNA that encodes A. vitis HR elicitor proteins. Use of
CC A. vitis HR elicitor proteins, or nucleic acids encoding them, may allow
CC control of previously untreatable diseases; provide systemic treatment;
CC and eliminate the need for biological control agents or polluting
CC chemicals.

XX Sequence 1668 BP; 441 A; 423 C; 462 G; 342 T; 0 other;

alignment_scores:

Quality: 123.50 Length: 209
Ratio: 1.056 Gaps: 6

Percent Similarity: 55.981 Percent Identity: 20.574

alignment_block:

US-09-466-935-2 x A61501/rev ..

Align seg 1/1 to reverse of: A61501 from: 1 to: 1668

10 LeuLeuThrSerIle.....IleLeuThrIleuSerProGlySerGlyAl 24
|||||
699 CTGATTCGTCGATTAACCTGCGTGCCTCTCAACCCCGCGGATTT 650
24 aIleAsnThrMetThrThrSerIleAsnHisGlyThrProAlaGly.... 39
649 TGGCATGATGTCGCGCAATAGTTTCTTATGCGCGCAAAATCCGCGTCG 600
40GlyValIlyrCysTPAlaSerAspArgThrGly 50
599 CGCGTCGATGGGCATTCGATTCGCTGCG..... 568
51 AspSerThrCysAlaGlyThrArgGlyValGlyThrLeuPheSerArg 67
567TTTCATGTGATCTATGCGCATGTTCGCGCATTCG 536
67 rValIle.....AlaPheGluValLeuThrPheArg 78
535 GATCATTCAGCATATTTCCCAATATCTGATTCATCAATATTCGTCG 486
78 lYAlaAlaThrLeuIleThrLeuGlyIle...GlnGlnTPArgAlaAla 93
485 GCGCGGCTATCTGTCGTCATGCGGCTTGGCCAGCGCTTCAGCAAGATA 436
94 GLYAlaIleAspLeuIleSerLeuAlaSerThrGlnSerArgAlaHisIle 110
435 CGCGATGTCGAGCGCATGCTGTTCCAGCGAGCATGAGCGCGCGGA 386
110 uPheGlnArgAlaValPheValAsnLeuThrAsnProIlySerIleValP 127
385 AATGATGACGGGTATCTGACCAATGTCATTCGCAAGACGTCGATTT 336
127 heLeuAlaAlaLeuPheProGlnIleMetProGlnIleProGlnIleu 143
335 TCGTGATGATGTCCTACACGCGCATTCATCGGACGAGCACGCCG...CTG 289
144 MetGlnThrIleValLeuGlyValThrThrIleValAlaPheIleIleVal 160
288 TCTCAACCAATTCCTTGGGCTGTCATTCCTCGCATTTGCTGTG 239
160 IMetIleGlyTYrAlaThrLeuAlaGlnArgIleAlaLeuThrIleLysG 177
238 GTTCGATTCGCTTCGACCTTTTGTCCAAACCGCGCATTCGACGTCG 189
177 lYrProLysGlnMetLysAlaLeuAsnLysIlePheGlySerLeuPheMet 193
188 TGTGTGTCGTCGACGCGCTGTCAATATCTGATTCGCGGTGTCGTCG 139
194 LeuValGlyAlaLeuLeuAlaSerAla 202
138 TCGCTGGGCGCATTCCTTTACCGCG 112

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT:A61502

seq_documentation_block:

ID A61502 standard; DNA: 1668 BP.

AC A61502;

XX 23-OCT-2000 (first entry)

DE A. vitis hypersensitive response elicitor-encoding DNA, SEQ ID NO:11.

XX Hypersensitive response elicitor protein; HR: disease resistance;
KW insecticide; fungicide; antiviral; bactericide; growth enhancer;
KW stress resistance; transgenic plant; ds.

DR MPI, 2000-414602/36.
 DR P-PSDB; Y99598.
 PT Novel *Escherichia* bacterium having enhanced L-threonine resistance due
 PT to enhanced RhtC protein activity, used to produce L-threonine,
 PT L-homoserine, L-valine and L-leucine -
 XX
 PS Claim 9; Page 14-15; 24pp; English.
 XX
 CC The present sequence is the L-threonine resistance gene, rhtC, from
 CC *Escherichia coli*. This sequence may be used to impart L-threonine
 CC resistance on *E. coli* bacteria, which would be useful for producing
 CC a high yield of L-threonine. L-threonine resistance means that the
 CC bacteria will be able to grow on a minimal medium containing
 CC L-threonine at a concentration at which the corresponding wild-type
 CC strain would not grow. Since the transformed bacteria can grow on the
 CC minimal medium, it can synthesise L-threonine, which accumulates. The
 CC accumulated amino acids can then be removed from the culture medium.
 CC The bacterium of the present invention may also be used to synthesise
 CC L-homoserine, L-valine and L-leucine at increased levels.
 CC
 SO Sequence 840 BP; 165 A; 192 C; 248 G; 235 T; 0 other;

alignment_scores:
 Quality: 122.00 Length: 219
 Ratio: 1.119 Gaps: 8
 Percent Similarity: 49.772 Percent Identity: 21.005

alignment_block:
 US-09-466-935-2 x A48443 ..

Align seg 1/1 to: A48443 from: 1 to: 840

```

11 LeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIleAsnTh 27
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
214 ATGTGACATGTTGGCGCTTATGAGCCCGGTCGGATTGCTTTTGTGT 263
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 rMetThrSerLeuAsnHisGlyTyrProAlaGly..... 39
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
264 CTCGTAGACCGCTGTCAGTCGTCCGTAAGAAGCAGATGAGGCGTGC 313
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
40 ..GlyValTyrCysTyrPAlaSerAspArgThrGlyAspSerTyrCysAla 55
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 TGGGCATTCCTCGCCGCTA.....ATG 336
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
56 GlyTPArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheG 72
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
337 GTTGGGCTGGGATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
72 vAlaLeuLysTrp.....AlaGlyAlaAlaTyrL 82
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 AAAAATGGCTGGCTGCATACGCTGATTANGTGGGCGGCTGCTATC 430
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
82 euILeTPrLeuGlyIleGlnGlnTrrPArgAlaAlaGlyAlaIleAsp 98
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
431 TCTGCTGATGCTTACCAATGCTACGT.....GGTGCACGCAAAAAA 474
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
99 LysSerLeuAlaSerThrGlnSerArgArgHisLeu..... 110
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
475 GAGGGGCTTCTGCACCTGCGCACAGAGTCGAGTGGCGAAAGTGAGG 524
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
111 ....PheGlnArgAlaValPheValAsnLeuThrAsnProLysSerIle 126
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
525 CAGTTTCTCGAAAGGTTTACGACCATTCGCAATCCGAAAGCAGATTA 574
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
126 alphaLeuAlaAlaLeuPheProGlnPheIleMetProGlnGlnPro 142
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
575 TCTACTTTGGCTCGGCTGTTTCATGTTGCGGTGATTAACGTTGGCACT 624
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
143 LeuMetGlnTyrIleValLeuGlyValThr..... 152
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
625 ACCGGGCGCTGGGCAATTTTGGCGCTGATGTCGAAGCGTGGCGTG 674

```

```

153 ....ThrIleValIleAspIleIleValIleMetIleGlyTyrAlaThrLeu 168
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
675 GTTACCGCTGCTGACAGCTGTTGCCCTCGCCCAATGCGCGGTGTT 724
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
168 lagIlnArgIleAlaLeuThrIleLysGlyProLysGlnMetLysAlaLeu 184
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
725 ATCAACGCTGCGGCAAGTGGATGATGCT..... 753
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
185 AsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuAlaLeu 201
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
754 .....TTTGGCGGCGGCTTATTGTCGCGATTGGCAATTGATTTAT 797
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 rAlaArg 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
798 TTGCGCG 804

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.A52691

seq_documentation_block:

ID A52691 standard; DNA: 636 BP.

AC A52691:

DT 03-JAN-2001 (first entry)
 XX
 XX
 DE *Escherichia coli* yga gene.
 XX
 XX
 KW *E. coli*; yga gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein; ds.
 XX
 XX
 OS *Escherichia coli*.
 XX

EH Key Location/Qualifiers
 FT CDS 1..636
 FT /*tag= a
 FT /product= "yga"

PN EP1016710-A2.

PD 05-JUL-2000.

PF 17-DEC-1999; 99EP-0125263.

PR 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

PA (AJIN) AJINOMOTO CO INC.

PI L'vivites VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;

PI Tokimakova IL;

DR MPI; 2000-414802/36.

DR P-PSDB; B01789.

PT increased production of L-amino acids by an *Escherichia* bacterium
 PT comprises increasing the expression amount of an L-amino acid excretion
 PT protein -
 PT

PS Disclosure: Page 24, 29pp; English.

XX
 CC The present sequence is the yga gene (an excretion protein gene) of
 CC *Escherichia coli*. The amino acid excretion protein produced from this
 CC gene is involved in the production of amino acids, and an increase in its
 CC expression leads to an increased accumulation of amino acids in the cell.
 CC In this case, an increase in arginine, glutamic acid and lysine is
 CC achieved if multiple copies of the gene are transfected into a bacterium.
 CC The bacterium used is *E. coli*.
 CC
 SO Sequence 636 BP; 120 A; 143 C; 180 G; 193 T; 0 other;

alignment_scores:

Quality: 113.00 Length: 209
Ratio: 0.926 Gaps: 10
Percent Similarity: 58.373 Percent Identity: 24.880

alignment_block:

US-09-466-935-2 x A52691 ..

Align seg 1/1 to: A52691 from: 1 to: 636

```

7 PheAlaTYR.....LeuLeuThrSerIleIleLeuTh 17
  |||:|||||
4 TTTCTATTACTTTCAGGCTTGCACTTGCGGCGGCTATATCCCTACC 53
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
  |||:||||| :|||:||||| :|||:|||||
54 GCTGCGTCACAAATCTTTGTG.....ATGANTCAGGCGCATACGTC 97
34 lSGlyTYRProAlaGly..GlyValTYrCysTrpAlaSerAspArgThr 49
  :|||:||||| :|||:||||| :|||:|||||
98 GTCAGTACACACATTTATGATTGCTTGTCTTGTCTATCTACAGCATTTGGTC 147
50 GlyAspSerTYrCysAlaGlyTYrPArgGlyValGlyThrLeuPheSerAr 66
  |||:||||| :|||:||||| :|||:|||||
148 .....CTGATTTGCCGCGGATTTTGGCGGACGCGCTTTATGATGCA 191
66 gSerValIleAlaPheGluValLeuYstrPAlaGlyAlaIleTYrLeuT 83
  :|||:||||| :|||:||||| :|||:|||||
192 GTCGCGCTGCTGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 241
83 lETrPleuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeuYs 99
  :|||:||||| :|||:||||| :|||:|||||
242 TGTGGTATGCTTTTGGCGCTTTTAAACACATGACGATGATTTATGAG 291
100 SerLeuAlaSerThr.....GlnSerArgArgHisLeuPheG1 112
  |||:||||| :|||:||||| :|||:|||||
292 ..TTAGCCAGCGCCGAGATCATGAAGCAAGCAGATGGAATATATATGCG 338
112 nArgAlaValPheValAsnLeuThrAsnPro.....LysSerI 125
  :|||:||||| :|||:||||| :|||:|||||
339 CACCATGTTGGCACTGACCTGCGCTGAATCCGATGTTTACTGATACTCT 388
125 lValIleAlaPheValAlaLeuPheProGlnPheIleMetProGlnInPro 141
  :|||:||||| :|||:||||| :|||:|||||
389 TTGTTGATCTGGCAGCTTGGCGGCA.....CTTATGTTGAACCA 432
142 GlnLeuMetGlnTYrIleValIleuGlyValThrThrIleValValAsp11 158
  :|||:||||| :|||:||||| :|||:|||||
433 AAA.....GCTGCTTTCACCTCGG.....ACATATAGCGCTCTTT 470
158 eIleValMetIleGlyTYrAlaThrIleuAlaGlnArgIleAlaLeuTrpI 175
  :|||:||||| :|||:||||| :|||:|||||
471 COTGCGTCTTTTGGCTGCTGCTCTCTCGAGGCTGCTGCGGACGCGCTG 520
175 lElySGlyProLYSGlnMetLYsAlaLeuAsnLYsIlePheGlySerLeu 191
  :|||:||||| :|||:||||| :|||:|||||
521 TGGCGACGCGCAAAAGCACAGCGCATATTCATCTGTGTGGATGTGT 570
192 PheMetLeuValGlyAlaLeuLeuAla 200
  :|||:||||| :|||:||||| :|||:|||||
571 ATGTGCTTATTTGCTTCGACGCTGCGG 597

seq_name: /STD2/gcgdata/geneseq/geneseq/NA197.DAT:T96816
seq_documentation_block:
ID T96816 standard; DNA; 2374 BP.
AC T96816;
XX
XX 12-MAR-1998 (first entry)
XX
XX DNA encoding LysG, LysE and ORF3 from Corynebacterium glutamicum.
XX
XX LysG; LysE; ORF3; Lysine transport; regulatory protein; export protein;

```

```

KW Microbial production; amino acid; animal feed additive; ds.
XX
XX Corynebacterium glutamicum.
XX
XX Key Location/Qualifiers
XX CDS complement (82..954)
XX CDS /tag= a
XX CDS /label= LysG
XX CDS 1016..1726
XX CDS /tag= b
XX CDS /label= LysE
XX CDS complement (1723..2373)
XX CDS /tag= c
XX CDS /label= orf3
XX
XX DE19548222-A1.
XX
XX 26-JUN-1997.
XX
XX 22-DEC-1995; 95DE-1048222.
XX
XX 22-DEC-1995; 95DE-1048222.
XX
XX (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
XX Eggeling L, Sahm H, VriJjc M;
XX
XX WPI; 1997-333867/31.
XX P-PSDB; W37714-16.
XX
XX Increasing microbial production of amino acids, especially lysine -
XX by improving export carrier activity or corresponding gene
XX expression, also new export and regulatory genes from
XX Corynebacterium
XX
XX Claim 23 and 26; page -: 16pp; German.
XX
XX This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
XX LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
XX protein and an export protein, respectively. Microbial production of
XX amino acids (A) is improved by increasing the export-carrier activity
XX and/or the export gene expression in a microorganism that produces (A).
XX The method is specifically used to increase production of lysine,
XX used as an animal feed additive. Other (A) are variously useful as
XX pharmaceuticals, condiments and intermediates for fine chemicals.
XX This method increases the amount of (A) secreted into the culture medium.
XX Export of (A) has been found to depend on a single gene.
XX NB. This sequence has been created from the information given in table 2
XX of the specification.
XX
XX Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other:

Alignment_scores:
Quality: 89.50 Length: 175
Ratio: 1.078 Gaps: 4
Percent Similarity: 47.429 Percent Identity: 20.571

alignment_block:
US-09-466-935-2 x T96816 ..

Align seg 1/1 to: T96816 from: 1 to: 2374

3 LeuGlnTrpTrpPheAlaTYrLeuLeuThrSerIleIleLeuThrLeuSe 19
  :|||:||||| :|||:||||| :|||:|||||
1025 ATGGAATCTTCATACAGCTCTTTTGGGCGCCAGCTCTTTACTGTC 1074
19 rProGlySerGlyAlaIleAsnThrMetThrThrSerLeu..AsnHisG 35
  :|||:||||| :|||:||||| :|||:|||||
1075 CATCGACGCGCAGATGTACTGTGATTAACAAGGATTAACCGGGAAG 1124
35 lTYrProAlaGlyGlyValTYrCysTrpAlaSerAspArgThrGlyAsp 51
  :|||:||||| :|||:||||| :|||:|||||

```

```

1125 GACTCATGCGGTTCTCTGTCGTTTAAATTCGACGCTTT..... 1168
52 SerTyrCysAlaGlyTrpArgIleValGlyThrLeuPheSerArgSerVa 68
   :::: ||||| ||||| ||| ::::
1169 TTGTTCATCCGCCGCGACCTTGSGCGTTGATCTTTTCCAAATGCCGCGC 1218
68 IleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTrpL 85
   ||||| :::: ||||| ||||| ||||| ::::
1219 GATCGTCCTCGATATTATGCGGTGGGTGCATCGCTTACCTGTTATGCT 1268
85 euGIYIleGIInGIInTrpArgIleAlaGlyAlaIleAspLeuLysSerLeu 101
   :::: :::: |||
1269 TTGCGCGTATGCGACGCAAGAGAGCC..... 1294
102 AlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValAlaPheValAs 118
1294 ..... 1294
118 nleuThrAsnProLysSerIleValPheLeuAlaAlaLeuPheProGlnP 135
   :::: ||||| :::: |||||
1295 .ATGACAAACAGAGGTGGAAGCG.....CCACAGA 1322
135 heIleMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuGlyVal 151
   :::: ||||| :::: |||||
1323 TCATTGAAGAAACAGAACACCGCTGCCGATGACACGCCCTTGGGCGGT 1372
152 ThrThrIleValIleValAspIleIleValMetIleGlyTyrAlaThrLeuAl 168
   :::: ||||| :::: |||||
1373 TCGCGGGTGGCGACTGACACGCGCAACGCGGTGGCGGTGAGTGAAGCT 1422
168 agIArgIleAlaLeuTrpIleLys 176
   :::: ||||| :::: |||||
1423 CGATTAACGACGCGGTTGGGTAAAG 1447

seq.name: /stds2/gcgcdata/geneseq/geneseq/NAL198.DAT:V27579
seq_documentation_block:
ID V27579 standard; DNA: 2289 BP.
XX
AC V27579;
XX
DT 26-OCT-1998 (first entry)
XX
DE Human hepatocyte nuclear factor 4 isoform alpha 4 DNA.
XX
KW Hepatocyte nuclear factor 4; HNF4 alpha 4; human;
KM non-insulin-dependent diabetes mellitus; NIDDM; therapy: diagnosis;
KM vaccine; ss.
XX
OS Homo sapiens.
XX
PN W09823780-A1.
XX
PD 04-JUN-1998.
XX
PF 25-NOV-1997; 97WO-0521614.
XX
PR 26-NOV-1996; 9605-0032043.
XX
PA (JOSL-) JOSLIN DIABETES CENT INC.
XX
PI Krolewski AS;
XX
DR WPI: 1998-322753/28.
XX
PT Use of hepatocyte nuclear factor 4 - for developing products for
PT diagnosing and determining risk of diabetes and for identifying
PT agents for treatment of diabetes
XX
PS Claim 11; Fig 3; 43pp; English.
XX
CC This nucleotide sequence encodes human hepatocyte nuclear factor
CC 4 isoform alpha 4 (HNF4 alpha 4). Sequences are also provided for

```

```

CC isoforms alpha 1 (see V27577), alpha 2 (see V27578) and gamma (see
CC V27580). HNF4 is a transcription factor whose ligand has not yet
CC been identified. The isoforms of HNF4 are reported to be derived
CC from alternative splicing events. The alpha isoforms have been
CC mapped to chromosome 20 and the gamma isoform to chromosome 8. A
CC claimed method for determining if an animal is at risk of
CC developing diabetes, comprises evaluating an aspect of HNF4
CC metabolism or structure in the animal, especially as related to
CC HNF4 from chromosome 20. An abnormality in the aspect of HNF4
CC metabolism or structure is diagnostic of being at risk for
CC diabetes, especially non-insulin-dependent diabetes (NIDDM). Also
CC claimed are: (1) a method for evaluating an agent for use in
CC treating diabetes by evaluating the effect of the agent on an
CC aspect of HNF4 metabolism; (2) the agent identified in (1); (3) a
CC method for treating a cell having an abnormality in structure or
CC metabolism of HNF4, comprising administering the agent to the cell;
CC (4) a method for producing a fragment or analogue of HNF4
CC polypeptide having wild-type activity; and (5) a method for
CC producing an HNF4 polypeptide or fragment or analogue having a
CC non-wild type activity.
XX
SQ Sequence 2289 BP; 489 A; 723 C; 653 G; 424 T; 0 other;

```

```

alignment_scores:
Quality: 85.50 Length: 172
Ratio: 1.043 Gaps: 7
Percent Similarity: 47.674 Percent Identity: 22.093

```

```

alignment_block:
US-09-466-935-2 x V27579/rev ..

```

```

Align seg 1/1 to reverse of: V27579 from: 1 to: 2289

```

```

18 LeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHi 34
   ||||| ||||| ||||| ::::
1471 CTGTCCCTGGGTGGGCGCCACTCACACATCTGCTTCGTGAGTGAG 1422
34 sGIYTrpProAlaGlyValIleTyrCysTrpAlaSerAspArgTrpGlyA 51
   |||
1421 TGGG..... 1418
51 spSerTyrCysAlaGlyTrpArgIleValGlyThrLeuPheSerArgSer 67
   :::: ||||| :::: |||||
1417 .....CATGTGTGGCAACGATGACGTGGTTC..... 1388
68 ValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTr 84
1387 .....CA 1386
84 pleuGIYIleGIInGIInTrpArgIleAlaGlyAlaIleAspLeuLysSerL 101
   :::: ||||| :::: |||||
1385 TATGTCTCTGATCAATGAGGTGAGGTGCGAGGGGTGGTCATGGGTCACA 1336
101 euAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValAlaPheVal 117
   :::: ||||| :::: |||||
1335 TCGCTGGGGACCTCCGACGACATCTCTCCACACAGTT.....GTC 1292
118 AsnLeuThrAsnProLysSerIleValPheLeuAlaAlaLeu.....Ph 132
   ||||| :::: ||||| :::: |||||
1291 AATCTTGGCCATGCCGAGAGCTTGATGACATCGATCTGTCATCATCT 1242
132 eProGlnPheIleMet.....ProGlnGlnProGlnLeuMetGln. 145
   :::: ||||| :::: |||||
1241 GCCACGTATGCTTCGCAAGGTGGGCAACGACGACGAGCTCTCCAAAG 1192
146 .....TyrIleValLeuGlyValThrThrIleValAlaValAspIle 159
   :::: ||||| :::: |||||
1191 CGGCCACGCGATCATACGCGGCTGATGATGATGATCTCCAAAGCTCAC 1142
160 ValMetIleGlyTyrAlaThrLeu.....AlaGlnArgIle 171
   :::: ||||| :::: |||||
1141 CTGACACTGGGAACGACGCCGCTGATCTTCCTCGATGAGCTCAGCCCT 1092

```

171 eAlaLeuTriLeuLys 176
 : : : : :
 1091 TGGCATCTGGGTCAA 1076

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA199.DAT:X13048

seq_documentation_block:
 ID X13048 standard; DNA: 6590 BP.

AC X13048;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:111.

KM Enterococcus faecalis; contig: detection; Enterococcal infection;

KW vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

PN W09850555-A2.

PD 12-NOV-1998.

PE 04-MAY-1998; 98WO-US08985.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046555.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT infection.

PS Claim 1; Page 700-703; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SO Sequence 6590 BP; 1981 A; 1397 C; 1108 G; 2091 T; 13 other;

alignment_scores:

Quality: 85.00

Ratio: 0.859

Percent Similarity: 41.597

Percent Identity: 18.908

alignment_block:

US-09-466-935-2 x X13048/rev ..

Align seg 1/1 to reverse of: X13048 from: 1 to: 6590

7 pheAlaTriLeuLeu.....ThreIleIleLeuTh 17

```

5143 TTTGCTTATTATTAACCTTTACACGTGGATTAACAGTGGCAGCGTTTGG 5094
17 rleuSerProglySerGlyAlaIle.....AsnThrMetThrThrs 31
5093 TGCACGCGCAGCGTGGTGGTGTATTTGGGCACTGCACATGTTAACAG 5044
31 erleuAsnHisGlyTyrProAlaGlyGlyValTyrCysTrpAlaSerasp 47
5043 GAATGAATCCGCGATGCTCCAAATCAGTAATATTTTC..... 5009
48 ArgThrGlyAspSerTyrCysAlaGlyTrpParGlyValGlyThrLeuph 64
5008 ...ACTGGGGAGACATTATTCGACAGACAGCGGAATT..... 4973
64 eSerArgSerValIleAlaPheGluValleuLysTrpAlaGlyAlaIat 81
4972 .....ATCGGTATTATT 4960
81 YrleuIleTrpleuGlyIleGingIntTrpArgAlaIaGlyAlaIleasp 97
4959 TTGCCGTATG..... 4949
98 leuLysSerleuAlaSerThrGlnSerArgHisleupheGlnArgAl 114
4948 TTATTATCATTTGTTAGAAAACAAATTGCATAAATGTGCCAGATCAT 4899
114 a...ValPheValAsnLeuThrAsnProLysSerIleValPheIleuAla 130
4898 TCATTTATTCGTAAACGCCAACATTTCACTTTAGTATTTGGCTTAAGCA 4849
130 leuPheProGlnPheIleMetPro..... 138
4848 CTAATTTTC.....TTAATCATGCTGTAGCCGAGCTATTCTTAATGAC 4805
139 .....GingInProGlnleuMetGlnTy 146
4804 TTAGTTGCTATTATCATGTTGCTTAGAAAAGGTGGCATGTTGCGG 4755
146 rIleValleuGlyValThr..... 152
4754 CTTTACTCTTGGTTTAACCTTGTGCCAATGTAATGTTGGTTTACATC 4705
153 .....Thr 153
4704 AAATTTGACGCCAATCCATTTGAAATGATTATTAACAAACAGGATGACT 4655
154 IleValIleAspIleIleValMetIleGlyTyrAlaThrleuAlaGlnAr 170
4654 TTGTTTATTCCAATTCTGGCATGCGTGGGCAAGTGGTGCAGC 4605
170 gIleAlaLeuTriPleuLysGlyProLysGlnMetLysAlaIleuAsnLysI 187
4604 TTTCATATTGATGATTCGTTGTAATCAGATAAATAATTACTTGAATGA 4555
187 lePheGlySerleu 191
4554 TTAAGGTGCTTG 4541

```

141 ProGlnLeuMetGlnTyrIleValLeuGlyValThrThrIleValValAs 1:


```

STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,219
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/029,679
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,056
FILING DATE: 02-OCT-1996
APPLICATION NUMBER: US 60/025,719
FILING DATE: 10-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 1441 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 20..1414
US-08-927-219-78

```

alignment_scores:

Quality:	82.50	Length:	172
Ratio:	1.019	Gaps:	7
Percent Similarity:	47.093	Percent Identity:	22.093

alignment_block:

US-09-466-935-2 x US-08-927-219-78/rev ..

Align seg 1/1 to reverse of: US-08-927-219-78 from: 1 to: 1441

```

18 LeuSerProGlySerGlyAlaIleAsnThrMetThrSerIleuAsnH1 34
|||||
1270 CTGTCCTCGGTCGGGCGCCACATCTGCTGCTGAGGTGAG 1221
34 SGLYIYrProAlaGlyValIYrCysTrpAlaSerAspArgThrGlyA 51
|||||
1220 TGGG.....
51 spSerTYrCysAlaGlyTrpArgIYValGlyThrIleuPheSerArgSer 67
:::|||||
1216 .....CATTTGTTGGCAGACATGACGTTGGTTC..... 1187
68 ValIleAlaPheGluValIleuLysTrpAlaGlyAlaAlaIYrIleuIeTr 84
1186 .....CA 1185
84 PleuGlyIleGlnInTrpArgAlaAlaGlyAlaIleAspIleuLysSerL 101
:::|||||
1184 TATGTTCTGCATGACGTGAGGAGTGCAGGGGTGGTGGGTGCA 1135
101 eAlaIleSerThrInSerArgArgHisIleuPheGlnArgAlaValPheVal 117
:::|||||
1134 TGGCTGGGGGACCTCCACGACGATCTCTGCAACAGGT.....GTC 1091

```

```

118 AsnIleThrAsnProLysSerIleValPheIleuAlaIleu.....Ph 132
|||||
1090 AATCTTGCCATGCCGAGAGCTGTGATGATCTGCTGATCARTCT 1041
132 eProGlnPheIleMet.....ProGlnInProGlnIleuMetGln 145
|||
1040 GCCAGGTGATGCTCTGCAGAGTGGGCGACGACGACGACGCTCTCCAAAG 991
146 .....TyrIleValIleuGlyValIYrThrIleValAlaIle 159
:::|||||
990 CGGCCAGCGAGTCACTAGCGGCTGTGATGATGATCTCCCAAGCTCAC 941
160 ValMetIleGlyTyrAlaIleu.....AlaGlnArgIle 171
:::|||||
940 CTGCACCTGGGAACGACGCCCTGTATCTCCCTGATGCTCAGCCCT 891
171 eAlaIleuTrpIleLys 176
:::|||||
890 TGGCATCTGGTCCAA 875

```

seq_name: /cgn2_6/plodata/2/ina/5B_COMB.seq:US-08-976-074-4

seq_documentation_block:

```

; Sequence 4, Application US/08976074
; Patent No. 5840847
; GENERAL INFORMATION:
; APPLICANT: ABRAMOVITZ, MARK
; APPLICANT: GRIGORCZYK, RICHARD
; APPLICANT: METTERS, KATHLEEN
; APPLICANT: NGUYEN, TRIYEN
; APPLICANT: RUSHMORE, THOMAS H.
; APPLICANT: SLIPETZ, DEBORAH
; TITLE OF INVENTION: DNA Encoding Prostaglandin Receptor FP
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. Mark Hand - Merck & Co., Inc.
; STREET: P.O. Box 2000 - 126 E. Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,074
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, J. MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19027DA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (732) 594-3905
; TELEFAX: (732) 594-4720
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-976-074-4

```

alignment_scores:

Quality:	79.50	Length:	211
Ratio:	0.710	Gaps:	12
Percent Similarity:	53.081	Percent Identity:	22.275

alignment_block:
US-09-466-935-2 x US-08-976-074-4 ..

Align seg 1/1 to: US-08-976-074-4 from: 1 to: 1437

```

21 GlySerGlyAlaIleAsnThrMetThrSerLeuAsnHisGlyTyrPr 37
   |||||
78 GCGAGCGCGCGCGCGGGG.....CGCCATGGCACACACC 109
   |||||
37 GAlaGlyValTyrCysTrpAlaSerAspArgThrGlyAspSerTyrC 54
   |||||
110 GAGGCGCTCCGCTTCTGC.....TCCACAGAGAGCCCGG 144
   |||||
54 ysalagly..TrpArggly...ValGlyThrLeupheserArgSerVal11 69
   |||||
145 CTGGCGGCGCTGGATGACAAAGATCTGTGACTGCATCTCCACAGTTT 194
   |||||
69 eAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTrpLeu 86
   |||||
195 GAGAGGAGATGACTGTAGTGG.....TTGGCTTTTATCT 229
   |||||
86 lytle.....GlnGlnTrpArgAlaGly 94
   |||||
230 CCACAAACATGTCATGACATTCACAAACGCTAGTGTCTCTCCAGACT 279
   |||||
95 AlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgHisLeuPh 111
   |||||
280 GCGCTTCTTCAACACACACCTGCCACAGCGAAACCGGCTTCCGATTT 329
   |||||
111 eGlnArgAlaValPheValAsn.....LeuThrAsnProLysSer1 125
   |||||
330 TTTTTCATGATCTTCATGACAGTGGGAAATCTGTCAACAGCCCTTCCA 379
   |||||
125 LeValPheLeuAlaIleAlaLeupheProGlnPheIleMetProGlnPro 141
   |||||
380 TCGCATCTTCATGACAGCATATCAGAGATT.....AGACAGAG 420
   |||||
142 GlnLeuMetGlnTyrIleValIleGlyValThrThrIleValValAsp1 158
   |||||
421 TCCAAAGCATCGTTTGTGCTTTGGCAGCGCGCTGTATATCATGATTT 470
   |||||
158 e.....IleValMetIleGlyTyrAlaThrL 167
   |||||
471 CTTTGGCCATCTCATCAATGAGCCATAGCATATTGTATATCTTCTG 520
   |||||
167 euAlaGlnArgIleAlaLeuTrpIleLysGlyProLysGlnMetLysAla 183
   |||||
521 ATTAAGAA.....TGGATCCGCTTT...GACCAATCAAAATGTC 555
   |||||
184 LeuAsnLysIlePheGlySerLeuPhe.....Me 193
   |||||
556 CTTGCGAGTATTTTGTATCTGCATGTGTCTTCTGTGTGCCACT 605
   |||||
193 tLeuValGlyAlaLeuLeuAlaSerAlaArg 203
   |||||
606 TCTTCTAGCGCATGTGTGATGGCCATTGAGCGG 636

```

seq_name: /csgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-083-741-4

seq_documentation_block:
Sequence 4, Application US/08083741
Patent No. 5869281
GENERAL INFORMATION:
APPLICANT: ABRAMOWITZ, MARK
APPLICANT: GRYGORCZYK, RICHARD
APPLICANT: METTERS, KATHLEEN
APPLICANT: NGUYEN, TUYEN
APPLICANT: RUSHMORE, THOMAS H.
APPLICANT: SLIPETZ, DEBORAH
TITLE OF INVENTION: DNA Encoding Prostaglandin Receptor FP
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: John W. Wallen, III

```

/ STREET: P. O. Box 2000, 126 E. Lincoln Avenue
/ CITY: Rahway
/ STATE: New Jersey
/ COUNTRY: USA
/ ZIP: 07065
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/083,741
/ FILING DATE: 25-JUN-1993
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: WALLEN, JOHN W.
/ REGISTRATION NUMBER: 35,403
/ REFERENCE/DOCKET NUMBER: 19027
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (908)594-3905
/ TELEFAX: (908)594-4720
/ INFORMATION FOR SEQ ID NO: 4:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1437 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: cDNA
/ US-08-083-741-4

```

alignment_scores:

Quality:	79.50	Length:	211
Ratio:	0.710	Gaps:	12
Percent Similarity:	53.081	Percent Identity:	22.275

alignment_block:
US-09-466-935-2 x US-08-083-741-4 ..

Align seg 1/1 to: US-08-083-741-4 from: 1 to: 1437

```

21 GlySerGlyAlaIleAsnThrMetThrSerLeuAsnHisGlyTyrPr 37
   |||||
78 GCGAGCGCGCGCGCGGGG.....CGCCATGGCACACACC 109
   |||||
37 GAlaGlyValTyrCysTrpAlaSerAspArgThrGlyAspSerTyrC 54
   |||||
110 GAGGCGCTCCGCTTCTGC.....TCCACAGAGAGCCCGG 144
   |||||
54 ysalagly..TrpArggly...ValGlyThrLeupheserArgSerVal11 69
   |||||
145 CTGGCGGCGCTGGATGACAAAGATCTGTGACTGCATCTCCACAGTTT 194
   |||||
69 eAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTrpLeu 86
   |||||
195 GAGAGGAGATGACTGTAGTGG.....TTGGCTTTTATCT 229
   |||||
86 lytle.....GlnGlnTrpArgAlaGly 94
   |||||
230 CCACAAACATGTCATGACATTCACAAACGCTAGTGTCTCTCCAGACT 279
   |||||
95 AlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgHisLeuPh 111
   |||||
280 GCGCTTCTTCAACACACACCTGCCACAGCGAAACCGGCTTCCGATTT 329
   |||||
111 eGlnArgAlaValPheValAsn.....LeuThrAsnProLysSer1 125
   |||||
330 TTTTTCATGATCTTCATGACAGTGGGAAATCTGTCAACAGCCCTTCCA 379
   |||||
125 LeValPheLeuAlaIleAlaLeupheProGlnPheIleMetProGlnPro 141
   |||||
380 TCGCATCTTCATGACAGCATATCAGAGATT.....AGACAGAG 420

```

```

142 GlnLeuMetClnTyrIleValLeuGlyValThrThrIleValAlaSpl1 158
      :::::::::::::::::::::
421 TCCAGGCAATCGTTCTGCTTTGGCCAGCGGCTGTATCATCATGATT 470
158 e:::::::::::::::::::::::::::llevalMetIleGlyTyrAlaThrL 167
      :::::::::::::::::::::
471 CTTTGGCCATCTCATCATGAGCCATAGCAGATTTGTATGCTTCG 520
167 euAlaGlnArgIleAlaLeuTyrIleLysGlyProLysGlnMetLysAla 183
      :::::::::::::::::::::
521 ATAAAGAA.....TGCATCCGCTTT...GACCAATCAATGTC 555
184 LeuAsnLysIlePheGlySerLeuPhe.....Me 193
      :::::::::::::::::::::
556 CTTTGCAATATTTTGTATCATGATGCTTTCTGCTGTGCCACT 605
193 tLeuValGlyAlaLeuLeuAlaSerLysArg 203
      :::::::::::::::::::::
606 TCTTCTAGGCAATGTGATGCGCATTGAGCGG 636

seq_name: /cgn2_6/ptodata/2/lna/5A_COMB.seq:US-08-764-233A-1

seq_documentation_block:
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neff, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:

```

```

; CLONE: p98/L, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "SorB"
; OTHER INFORMATION: /note= "this gene encodes a protein that is highly homolog
; OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs t
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product= "Sorb"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product= "Module 1 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product= "Module 2 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30881..35446
; OTHER INFORMATION: /product= "Module 3 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product= "Module 4 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40190..46318
; OTHER INFORMATION: /product= "Module 5 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "The protein encoded by the sorB gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygrosopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
; US-08-764-233A-1

```

```

alignment_scores:
      Quality: 79.50      Length: 103
      Ratio: 1.445      Gaps: 5
Percent Similarity: 53.398      Percent Identity: 32.039

```

```

alignment_block:
US-09-466-935-2 x US-08-764-233A-1 ..

```

```

Align seq 1/1 to: US-08-764-233A-1 from: 1 to: 49377

```



```
166 ThrLeuAlaGlnArgIleAlaLeuTrpIleAlaLeuTyrProIysGlnMetIly 182
      |||:::
771 ATGCTTGTTCA.....TTTAAAGAAATTAATGAGGTGATTGACAGA 814
      |||:::
182 sAlaLeuAsnIlyIlePheGlySerIleuPheMet 193
      |||:::
815 AGCATTCTCTTCTTCTGCTCAGTCAGTATTCATG 848

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-927-219-130

seq_documentation_block:
; Sequence 130 Application US/08927219
; Patent No. 6187353
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamagata, Kazuya
; APPLICANT: Oda, Naohisa
; APPLICANT: Katsaki, Pamela J.
; APPLICANT: Furuta, Hiroto
; APPLICANT: Horikawa, Yukio
; APPLICANT: Menzel, Stephen
; TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
; TITLE OF INVENTION: GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
; TITLE OF INVENTION: AND HNF-4ALPHA
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/927,219
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/029,679
; FILING DATE: 30-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,056
; FILING DATE: 02-OCT-1996
; APPLICATION DATA:
; APPLICATION NUMBER: US 60/025,719
; FILING DATE: 10-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:272
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 512/418-3000
; TELEFAX: 512/474-7577
; INFORMATION FOR SEQ ID NO: 130:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10014 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-927-219-130

alignment_scores:
      Quality: 78.00      Length: 194
      Ratio: 0.867      Gaps: 10
      Percent Similarity: 46.392      Percent Identity: 22.680

alignment_block:
US-09-466-935-2 x US-08-927-219-130  ..
```

```
Align seq 1/1 to: US-08-927-219-130 from: 1 to: 10014
14 lleileuThrIleuSerProGlySerGlyAlaIleAsnThrMetIly 30
      :::::|||||:::
5619 AVGTTGCTTACAGTTTTCATCAGCAGACAGAAAGGCCAGCCAGACGA 5668
      |||:::
30 rSerLeuAsnHisGlyTyrProAlaGlyGlyValTyrCysTrpIleSerA 47
      |||:::
5669 GTTCTTGCCCAAGGACAC.....AGCAGTT 5694
      |||:::
47 sPArgThrGlyAspSerTyrCysAlaGlyTrpArgGlyValGlyThrIleu 63
      ::|||:::
5695 CAGGACAGAGATCAGC.....CCAGGCTCTCAGCTCTG 5729
      |||:::
64 PheSerArgSerValIleAlaPheGluValLeuIlyStrpAlaGlyAlaI 80
      |||:::
5730 ..AGCACATGTTCTTCTCCCTTCAGGTTTCTAGTTTATGGGTAGTAG 5776
      ::|||:::
80 aTyrIleuIleTrpLeuGlyIleGlnIleTrpArg.....AlaAlaG 94
      ::|||:::
5777 TTTTATG.....ATGCCATTTCAGCTTCAGGAGGTAGAGCAGCAGG 5820
      |||:::
94 lYAlaIleAspLeuIlySerLeuAlaSerThrGlnSerArgIlyHisIleu 110
      |||:::
5821 GGAGCATT..... 5828
      |||:::
111 PheGlnArgAlaValAlaPheValAsnLeuThrAsnProIysSerIleValPh 127
      ::|||:::
5829 .....AMCTGACTTGCCCAAGGCTCAGGCTCAGTGTG 5857
      |||:::
127 eLeuAlaIleuPheProGlnPheIleMetProGln..ProGlnL 143
      ::|||:::
5858 GTTACGGGAGGCTTCCCAAGGCTACAGATGCCAAACACTGTCTTCTTC 5907
      |||:::
143 eumetGln.....TyrIleValIleu 149
      |||:::
5908 TCTTTCAGGTGCTGCTGCTCAGAGCCATGCTGCGAGACACTGCTGCTC 5957
      |||:::
150 GlyValThr.....ThrIleValIleValIleValMetIleGlyTyr 164
      |||:::
5958 GGAGCCACCAAGAGATCCATGTTTCAAGACAGTGTCTCTAGGTGA 6007
      |||:::
164 rAla..... 165
      |||:::
6008 GCGCGTGCCTGCTGCTGCGCAGGCTCCAGGAGGTATGCTTACATGG 6057
      |||:::
166 .....ThrLeuAlaGlnArgIleAlaLeuTrp 174
      |||:::
6058 CACTCACCCAGCAAGAGATTCACATGTTG 6089

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-444-644-27

seq_documentation_block:
; Sequence 27 Application US/08444644
; Patent No. 6015555
; GENERAL INFORMATION:
; APPLICANT: Friden, Phillip M.
; TITLE OF INVENTION: TRANSFERIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTI-BODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,644
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,246
FILING DATE: 07-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,458
FILING DATE: 26-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US90/05077
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/404,089
FILING DATE: 07-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Wagner, Richard W.
REGISTRATION NUMBER: 34,480
REFERENCE/DOCKET NUMBER: ALK88-15AAAZ
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 10785 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
IMMEDIATE SOURCE:
CLONE: PAH4625
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1..10785
OTHER INFORMATION: /note="Function = "Expression
OTHER INFORMATION: Vector Coding Sequence""
US-08-444-644-27

alignment_scores:
Quality: 77.50 Length: 169
Ratio: 0.969 Gaps: 9
Percent Similarity: 47.337 Percent Identity: 25.444

alignment_block:
US-09-466-935-2 x US-08-444-644-27 ..

Align seg 1/1 to: US-08-444-644-27 from: 1 to: 10785

21 gLySerGIyAlAlleAsnThrMetThrThSer..... 31
|| :||:|||||:|||||:|||||:|||||:|||||: 31
6260 GGAGTGGCCAGCATTAATACCGTTGAGAGAGCGGGCGGACCCGACG 6309
32 ...LeuAhnHsGIyTyTrpAlaGlyValTyTyCysTrp..... 44
|| :||:|||||:|||||:|||||:|||||:||||| 44
6310 AATATACAAACCCAACTGAAGCTACGCGCGTGAACCTGCGCAACGA 6359
45 .....AlaSerAspArgThrGlyAspSerTyrcysAlaIyTrp 57
6360 CGCGTGGCTGGCTTTTACATCTACCGCGCGCATTTTGGCGCACTGTGG 6409
58 ArgGIyValGIyThrLeuPheSerArgSerValIleAlaPheGluValle 74
|| :||:|||||:|||||:|||||:|||||:||||| 74
6410 AAG...GTTCAATATTTCTTACTGCGACCGCATCGCC..... 6445
74 uysTrpAlaGlyAlaAlaTyrlleuIleTrpLeuGlyIleGlnGlnTrpA 91
6445 ..... 6445
91 rGAlaAlaGlyAlaIleAsnLeuTySeriLeuAlaSerThrGlnSerArg 107
|| :||:|||||:|||||:|||||:|||||:||||| 107
```

```

seq_name: /cgn2_6/prodata/2/lna/6A_COMB.seq:US-08-444-644-41
seq_documentation_block:
Sequence 41, Application US/08444644
Patent No. 6015555
GENERAL INFORMATION:
APPLICANT: Fiden, Phillip M.
TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
TITLE OF INVENTION: ANTIBODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millita Drive
CITY: Lexington
STATE: MA
COUNTRY: USA
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,644
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/232,246
FILING DATE: 07-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/800,458
FILING DATE: 26-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US90/05077
FILING DATE: 07-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/404,089
FILING DATE: 07-SEP-1989
ATTORNEY/AGENT INFORMATION:
NAME: Wagner, Richard W.
REGISTRATION NUMBER: 34,480
REFERENCE/DOCKET NUMBER: ALK88-15AAAZ
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 10844 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
```



```
32 ....LeuSnhISgLYrProAlaGlyValTyrCysTrp..... 44
      |||
      :|||
9114 AATATACAACTGAGCTACCGAGCGGCTGCGAAGCA 9163
45 .....AlaserparGhrGlyAspSerTyrCysAlaGlyTrp 57
      :|||
9164 CGCCGCGCTGCTTCTACATCTACCGGCGGAGTCTTGGCGGTGTG 9213
58 ArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValIle 74
      :|||
9214 AAG...GTTTCATATCTTCTTACGCCACCGCATCGCC..... 9249
74 uLysTrpAlaGlyAlaValTyrLeuIleTyrPLeuGlyIleGlnGlnTrpA 91
9249 ..... 9249
91 rGAlaAlaGlyAlaIleAspLeuYsSerLeuAlaSerThrGlnSerArg 107
      :|||
9250 ..TGTTTACGCTGCTGCTCAGCGGCTGCGGCGGCGGATCTCTCT 9297
108 ..ArgHISLeuPheGlnArgAlaValAlaPheValAsnLeuThrAsnProLys 123
      :|||
9298 GCAGACACCTT...CAGCGC.....GTAACCGT 9323
124 SerIleValPheIleuAlaAlaLeuPheProGlnPheIleMet.ProGlnG 140
      :|||
9324 GGTATTATCAAACTTCCGCTGTATCCCGACGAGGCTCATTCGCGCGGT 9372
140 InProGlnLeuMetGlnTyrIleValLeuGlyValThrThrIleValVal 156
9373 .....GCTTTCAGGTTATCGAGAAATATCGTTA 9399
157 AspIleIleValMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaIle 173
      :|||
9400 ACAGTCGGGGTAATGCTTTCAGAGGCGGAATATGCGCGGCGCTTAACAG 9449
173 uTrp 174
      |||
9450 CTGG 9453

seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-444-644-32
seq_documentation_block:
; Sequence 32, Application US/08444644
; Patent No. 6015555
; GENERAL INFORMATION:
; APPLICANT: Fiden, Phillip M.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTI-BODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,644
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,246
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,458
; FILING DATE: 26-NOV-1991
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/05077
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/404,089
; FILING DATE: 07-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Wagner, Richard W.
; REGISTRATION NUMBER: 34,480
; REFERENCE/DOCKET NUMBER: ALK88-15AAAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12127 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: pAH4807
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..12127
; OTHER INFORMATION: /note= "Expression
; OTHER INFORMATION: Vector Coding Sequence"
US-08-444-644-32
```

alignment_scores:

Quality:	77.50	Length:	169
Ratio:	0.969	Gaps:	9
Percent Similarity:	47.337	Percent Identity:	25.444

alignment_block:

US-09-466-935-2 x US-08-444-644-32 ..

Align seg 1/1 to: US-08-444-644-32 from: 1 to: 12127

```
21 GlySerGlyAlaIleAsnThrMetThrSer..... 31
      |||
7602 GGAGTCGCCACGATTAATATACCGTTGAGAAAGCGGCGGAGCCGCAAG 7651
32 ....LeuSnhISgLYrProAlaGlyValTyrCysTrp..... 44
      |||
7652 AATATACAACTGAGCTACCGAGCGGCTGACCTGCGGCAAGCA 7701
45 .....AlaserparGhrGlyAspSerTyrCysAlaGlyTrp 57
      :|||
7702 CGCCGCGCTGCTTCTACATCTACCGGCGGCACTTTTGGCGAGTGTG 7751
58 ArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValIle 74
      :|||
7752 AAG...GTTTCATATCTTCTTACGACCGCATCGCC..... 7787
74 uLysTrpAlaGlyAlaValTyrLeuIleTyrPLeuGlyIleGlnGlnTrpA 91
7787 ..... 7787
91 rGAlaAlaGlyAlaIleAspLeuYsSerLeuAlaSerThrGlnSerArg 107
      :|||
7788 ..TGTTTACGCTGCTGCTCAGCGGCTGCGGCGGCGGATCTCTCT 7835
108 ..ArgHISLeuPheGlnArgAlaValAlaPheValAsnLeuThrAsnProLys 123
      :|||
7836 GCAGACACCTT...CAGCGC.....GTAACCGT 7861
124 SerIleValPheIleuAlaAlaLeuPheProGlnPheIleMet.ProGlnG 140
      :|||
7862 GGTATTATCAAACTTCCGCTGTATCCCGACGAGGCTCATTCGCGCGGT 7910
140 InProGlnLeuMetGlnTyrIleValLeuGlyValThrThrIleValVal 156
```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1122)
US-08-878-801-3
```

```
alignment_scores:
    Quality: 76.50      Length: 209
    Ratio: 0.900        Gaps: 12
    Percent Similarity: 40.670    Percent Identity: 23.445
```

alignment_block:

US-09-466-935-2 x US-08-878-801-3/rev ..

Align seg 1/1 to reverse of: US-08-878-801-3 from: 1 to: 1125

```
5 TTTPTTP.....PhealatyrlleuethrSer11 14
|||||
568 TGGTGGCATGCGACTGCGCAGCAGCTCTTGGCGAGTGGGAGTACCTG 519
14 elleuethrleuSerProglYSerGlyAlaIleasnhrMetThrphs 31
|||||
518 TCCTCTGATATGCGCTCCAGGT.....GTGACAG 490
31 erleuasnhrleuSerProglYSerGlyAlaIleasnhrMetThrphs 47
|||||
489 GTAAATACCGCGCGAGATCGAGAGGTGGAATTCACGCTTCGCTGACG 440
48 .....ArgthrGlyaspSerTyrr..... 53
439 ATGCACGATGCCCGCGTCCGCCACAGAGTACTGATGCGACATGCAAT 390
54 .....CysAlaGlyTrpArgGlyValGlyThrleuPheSerA 66
|||||
389 GGCTTCTCGAATGTGCTCACTTTATAGGGGTCCTGGTCA..... 350
66 rgservalllealaphhegluvalleuSerTyrr..... 76
349 .....TCACTAGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 320
77 .....AlaGlyAlaAlaTyrrleuIleuTyrrleuG 86
|||||
319 GCCTGCTGAAGGGAGATCTGACAGCGGTCCATCGCATATCATGATG.... 275
86 yllieglnglnTrpArg.....AlaAlaGly 94
|||||
274 .....CCTGCATGAGACGAGATGTTCTGTGTAGATGACAGCCGAGAGC 229
95 AlaIleaspLeuIleuSerleuAlaSerThrGlnSerArgArgHisleuPh 111
|||||
228 TCTGCGGTCTCTCTCGAGTACCCACACCGCATGATGCGCATCTGCT 179
111 eglarGAlaValPheValAsnleuThrAsnPro.....LysSerIleV 126
|||||
178 TGATGACAGTACTCTCCGCTCTCAGCAGGCGCCACAGCAGGAGTTTC 129
126 alphaIleAlaAlaIleuPheProGlnPheIleMetProGlnGlnProGln 142
|||||
128 AATTCTCGCGCTCTGTTT.....TTTCTGTCCAAACCAATCT... 88
143 leuMetGlnTyrrIleValleuGlyValThrThrIleValAlaIleIle 159
|||||
87 .....GTTGATCTTCCT 77
159 eValMetIleGlyTyrrAlaThrleu 167
|||||
76 GGTGATTCCTGGCGCAGTCTTCTC 52
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-453-695A-113
seq_documentation_block:
```

```
; Sequence 113, Application US/08453695A
; Patent No. 5708143
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: Protocadherin Materials and Methods
; NUMBER OF SEQUENCES: 115
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/453,695A
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5708143and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 32658
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ. ID NO: 113:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2347 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-453-695A-113
```

```
alignment_scores:
    Quality: 75.50      Length: 76
    Ratio: 1.606        Gaps: 4
    Percent Similarity: 61.842    Percent Identity: 32.895
```

alignment_block:

US-09-466-935-2 x US-08-453-695A-113 ..

Align seg 1/1 to: US-08-453-695A-113 from: 1 to: 2347

```
5 TTTPTTPheAlaTyrrleu.....leuThrSerIleIleuThrle 18
|||||
1189 TGGTGGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 1238
18 user: ProglYSerGlyAlaIleasnhrMetThrThrSerleuAsnHis 34
|||||
1239 GTCCTCTAGCTCTCAGGCT.....TCAGCCTGGCAGAA 1273
35 GlyTyrrProAlaGlyValTyrrCysTrpAlaSerAspArgThrGlyAs 51
|||||
1274 GTCCTGCCAGCTCCAGGGCTCCAAAGGTGATGTGGCTCTCGACCTGCAT 1323
51 pSerTyrrCysAlaGlyTrpArgGlyValGly.....ThrleuPhe 65
|||||
1324 CACATTTGTGGGCATCGACGGGTACAGGCTTTTCTACAAACCTATTCT 1373
65 erArgServalllealaphhegluval 73
|||||
1374 CATGAAGTCTGCTCACTTCAGGCTC 1399
```

9

1

2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:34:19 ; Search time 26.05 Seconds
(without alignments)
540.814 Million cell updates/sec

Title: US-09-466-935-2

Perfect score: 1061

Sequence: 1 MTEWMFAYLITSTILLSP.....KIRGSLFMLVGLLASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR 67:*

- 1: pir1:*
- 2: pir2:*
- 3: pir3:*
- 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	686	64.7	138	2	A65187	hypothetical 15.4
2	459.5	43.3	205	2	G82358	conserved hypothet
3	389	36.7	209	2	C82990	hypothetical prote
4	228	21.5	241	2	C75329	conserved hypothet
5	195.5	18.4	212	2	F64940	hypothetical prote
6	195	18.4	210	2	F69975	dihydrodipicolinat
7	193	18.2	210	2	G83082	hypothetical prote
8	192.5	18.1	207	2	E83703	hypothetical prote
9	190	17.9	212	2	G82200	conserved hypothet
10	186	17.5	208	2	C82471	conserved hypothet
11	180.5	17.0	207	2	D83187	hypothetical prote
12	180.5	17.0	216	2	F83051	conserved hypothet
13	178.5	16.8	213	2	C82523	amino acid transpo
14	178.5	16.8	216	2	C85580	hypothetical prote
15	166	15.6	206	2	C82139	conserved hypothet
16	156	14.7	204	2	B83729	hypothetical prote
17	155	14.6	153	2	S23222	hypothetical prote
18	154.5	14.6	208	2	G84086	miga protein - She
19	150	14.1	204	2	F83306	hypothetical prote
20	147	13.9	206	2	S76178	hypothetical prote
21	146	13.8	213	2	F83444	hypothetical prote
22	143.5	13.5	204	2	B82410	conserved hypothet
23	141	13.3	211	2	T36373	hypothetical prote
24	132	12.4	209	2	E82388	conserved hypothet
25	127.5	12.0	222	2	F82353	yifub protein (limp
26	126.5	11.9	212	2	T43921	membrane protein y
27	126.5	11.9	223	2	H64759	hypothetical 21.2
28	126	11.9	195	2	A65036	hypothetical prote
29	124	11.7	210	2	D84016	

30	123.5	11.6	197	2	B83280	hypothetical prote
31	116	10.9	203	2	T30317	chemotaxis protein
32	113	10.7	211	1	Q0EC5A	hypothetical 23k p
33	109	10.3	205	2	B83356	conserved hypothet
34	101.5	9.6	4588	2	T28667	dyeine beta heavy
35	94	8.9	457	2	B64769	proline transport
36	92	8.7	449	2	F83328	probable sodium/al
37	89.5	8.4	366	2	F83970	stage V sporulatio
38	88	8.3	204	2	B82166	conserved hypothet
39	87.5	8.2	465	2	A83046	probable amino aci
40	87	8.2	210	2	F64115	hypothetical prote
41	84.5	8.0	200	2	D83100	probable transport
42	84	7.9	190	2	T03459	hypothetical prote
43	84	7.9	297	2	H83888	transmembrane lipo
44	83	7.8	674	2	E82954	conserved hypothet
45	82.5	7.8	149	2	C83056	hypothetical prote

ALIGNMENTS

RESULT 1
A65187
hypothetical 15.4 kD protein in recQ-PlDb intergenic region - Escherichia coli (strat
N:Alternate names: hypothetical protein fl38
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: A65187; S30714
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.: Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A65187
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-138 <BIAT>
A:Cross-references: GB:AE000458; GB:U00096; NID:q267299; PIRN:AACT6827.1; PID:q17902
R:Experimental source: strain K-12, substrain MG1655
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.
A:Reference number: S30660; MUID:92358234
A:Accession: S30714
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'V', 2-138 <DAN>
A:Cross-references: EMBL:M87049
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: y1gk

Query Match 64.7% Score 686; DB 2; Length 138;
Best Local Similarity 99.3%; Pred. No. 5.7e-58;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAPFVLKAGAAAYLIWIGIOQWRAAGAIIDLKSLASTOSRRHLFORAVNLTNPKSIVF 127
: |||||||
Db 1 MAIFVLKAGAAAYLIWIGIOQWRAAGAIIDLKSLASTOSRRHLFORAVNLTNPKSIVF 60
: |||||||

QY 128 IAAFPQFIMPOQPOLMOYIVLGVTTIVDITWIGVTTAAQRIALWIKGKOKKALNKI 187
: |||||||
Db 61 IAAFPQFIMPOQPOLMOYIVLGVTTIVDITWIGVTTAAQRIALWIKGKOKKALNKI 120
: |||||||

QY 188 FGSFLFMLVGLLASARHA 205
: |||||||
Db 121 FGSFLFMLVGLLASARHA 138
: |||||||

RESULT 2
G82358
conserved hypothetical protein VC0136 [imported] - Vibrio cholerae (group O1 strain N


```
OY 1 MTELEWFAVLTSLITLSPGSAINTMTSLNHGYPAGGYCMASDRTGDSYACAGR--- 58
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MDLTSLSLFAGVAVLTLPMPGPILTVLAQSMQNRQAGV-----TALGCTGTLVH 53
OY 59 -----GVGTLFSRSYIAFEVLKMGAAVLTLMGIQOMRAA---AIDLKSLASTOSRRH 109
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 ITAAVVGISALTIYQSALATFTVYKAAAYILYLAAMKAFQEKGGSLSDKO---TTLAYGA 110
OY 110 LFORAVFVNLTPNKSIVFLAALFPQFIMPOQPOL-MQYIVLGVTTIVVDIIVMIGVATLA 168
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 LYKKGJIMVNLNPKVSLFELALLPQFVNSGAGSAPMOMLLGVFLIQAFILPSVSWFA 170
OY 169 QRIA-LMIGPKMKALNKIFGSLFMLVGALLASARH 204
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 EKVGLLMRSSPFIKNOMHRIKGLLALIGLVAFSKN 207

RESULT 9
G82200
conserved hypothetical protein VC1421 [imported] - Vibrio cholerae (group O1 strain N169
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: G82200
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: G82200
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <HEI>
A:Cross-references: GB:AE004221; GB:AE003852; NID:99655918; PIDN:AAF94578.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1421
A:Map position: 1

Query Match 17.9%; Score 190; DB 2; Length 212;
Best Local Similarity 27.4%; Pred. No. 8.9e-11;
Matches 58; Conservative 35; Mismatches 99; Indels 20; Gaps 5;

OY 1 MTELEWFAVLTSLITLSPGSAINTMTSLNHGYPAGGYCMASDRTGDSYACAGW--- 57
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MVIQNEFAFFIATITLTPGDLTALVIRNTSRAGFADG---C---TSLGICGGLFVH 53
OY 58 -----RGVGTLSRSYIAFEVLKMGAAVLTLMGIQOMRA---AGAILDKSLASTOSRR 108
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 54 APTSAIGISAILIAQSAEFLQIVKMGAAVLTLMGISLSRLMKKTGGGIEVASIAHAOFPL 113
OY 109 HLFORAVFVNLTPNKSIVFLAALFPQFIMPOQPOLMOMIVLGVTTIVVDIIVMIGVATL 167
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 114 TSLRGEFLSNVNLNPKVFLAFLPQFIPNIPYSPLAQSLMALIHFAIYMWQCGLAGA 173
OY 168 AQRIALMWIKPKOMKALNKIFGSLFMLVGALL 199
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 USSAKNLKLNASFMFMMEGTGTGVLVALGIKL 205

RESULT 10
G82471
conserved hypothetical protein VCA0355 [imported] - Vibrio cholerae (group O1 strain N16
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: G82471
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
```

```
A:Accession: G82471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-208 <HEI>
A:Cross-references: GB:AE004372; GB:AE003853; NID:99657741; PIDN:AAF96263.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0355
A:Map position: 2

Query Match 17.5%; Score 186; DB 2; Length 208;
Best Local Similarity 29.3%; Pred. No. 2.1e-10;
Matches 63; Conservative 35; Mismatches 89; Indels 28; Gaps 9;

OY 1 MTELEWFAVLTSLITLSPGSAI---NTMTSLNHGYPAG-GYYCMASDRTGDSY-- 53
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MDLNSILFIVACIAINMIPGPDVITYVNTMKGLVTGKRAMGL-----GVGYEV 52
OY 54 --CAGNRGVGTLSRSYIAFEVLKMGAAVLTLMGIQ---QWRAGAIDLKSLASTOSR 107
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 53 HTLLASLIGLSATILSSAVAFSAVKMIGAAVLYLVGQSLSMWR--GGSILKYSSEVED 110
OY 108 RHLFORAVFVNLTPNKSIVFLAALFPQFIMPOQPOL-MQYIVLGVTTIVVDIIVMIGVAT 166
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 KNVEVGVIVSVLNPKEVALLFSLFQFIDTSSGSASQMLWLGLFSLVLTMCNIIYAS 170
OY 167 LAQRIALMWIKPKOMKALNKIFGSLFMLVGALLAS 201
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 171 VGSWVFSRPNRSORYSRGLGCVGL--LIG--LAS 201
```

```
RESULT 11
D83187
hypothetical protein PA3665 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83187
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: D83187
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <STO>
A:Cross-references: GB:AE004786; GB:AE004091; NID:99949822; PIDN:AA607053.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3665
C:Superfamily: hypothetical protein b1798

Query Match 17.0%; Score 180.5; DB 2; Length 207;
Best Local Similarity 26.5%; Pred. No. 6.9e-10;
Matches 57; Conservative 34; Mismatches 97; Indels 27; Gaps 6;
```

```
OY 6 WFAVLTSLITLSPGSAINTMTSLNHGYP-----AGGYCMASDRTGDSYACAGW---- 57
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 5 WALFVPACFALNLAGPNNLTLNNAARHGFAVTASLAGGRLLA-----FAGMLALA 56
OY 58 -RGVGTLSRSYIAFEVLKMGAAVLTLMGIQOMRAAGAILDKSLASTOSRR-----HLF 111
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 ASGLALVHTSTWMLFLAIKLGAAVLTLMVAYOLWPT---DAQPLANEASPARPSIMLRG 112
OY 112 QRAVAVNLTPNKSIVFLAALFPQFIMPOQPOLMOMIVLGVTTIVVDIIVMIGVATLAORI 171
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 113 RQEFIVAAQGNPRAILIFALPQFIPDPQPLGAGPAGQAGAAFLLEWLAIALYVAGLHL 172
OY 172 ALMWIKPKOMKALNKIFGSLFMLVG-ALLASARRA 205
| | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:38:09 ; Search time 16.79 Seconds
(without alignments) 418.247 Million cell updates/sec

Title: US-09-466-935-2
1061

Perfect score: 1 MTELEWFAVLLTSLITLSP.....KIRGSLFMLVGLALASARRA 205
Sequence:

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	898.5	84.7	206	1 RHTB_ECOLI	P27847 escherichia
2	195.5	18.4	212	1 YEAS_ECOLI	P76249 escherichia
3	195	18.4	210	1 YRHP_BACSU	O05406 bacillus su
4	180.5	17.0	216	1 YBF7_PSEAE	P38102 pseudomonas
5	147	13.9	206	1 YG27_SYN3	P74343 synechocyst
6	126.5	11.9	223	1 YAHN_ECOLI	P75693 escherichia
7	126	11.9	195	1 YFIR_ECOLI	P38101 escherichia
8	122	11.5	206	1 RHTC_ECOLI	P27846 escherichia
9	116	10.9	203	1 CHPE_PSEAE	O87005 pseudomonas
10	113	10.7	211	1 YGCA_ECOLI	P11667 escherichia
11	111.5	10.5	206	1 YGSA_AERSA	P70775 aeromonas s
12	94	8.9	457	1 PROY_ECOLI	P77327 escherichia
13	92	8.7	456	1 PROY_SALTY	P37460 salmonella
14	89.5	8.4	236	1 LYSE_CORGL	P94633 corynobacte
15	87	8.2	210	1 YD07_HAEIN	O57320 haemophilus
16	81	7.6	732	1 ATZN_ECOLI	P37617 escherichia
17	79.5	7.5	384	1 CYNA_ECOLI	P17583 escherichia
18	79.5	7.5	589	1 FUR4_SCHPO	Q10279 schizosach
19	79	7.4	579	1 YN05_CAEEL	Q03614 caenorhabdi
20	78.5	7.4	366	1 SPSE_BACSU	P07373 bacillus su
21	78.5	7.4	471	1 PROX_MYXXA	P56601 myxococcus
22	78	7.4	1694	1 SN_MOUSE	O62230 mus musculu
23	77.5	7.3	225	1 YGGA_AERHY	P52047 aeromonas h
24	77	7.3	348	1 OPSD_SARXA	P79914 sarcocentro
25	77	7.3	543	1 SGLT_VIBPA	P96169 vibrio para
26	75.5	7.1	261	1 SP31_BACSU	O01625 bacillus su
27	75.5	7.1	445	1 IF36_HUMAN	O64452 homo sapien
28	75.5	7.1	460	1 AKCD_ECOLI	P77429 escherichia
29	75.5	7.1	566	1 KHT2_KJOLA	P53387 kluyveromyc
30	75	7.1	373	1 Y38K_THETE	P30715 thermoprote
31	75	7.1	440	1 YJ99_MYCTU	Q10858 mycobacteri
32	74.5	7.0	394	1 FT5W_HAEIN	P45064 haemophilus
33	74.5	7.0	460	1 NU4M_ONCMY	P11631 oncoirynchu

34	74.5	7.0	504	1 OCLN_CHICK	Q91049 gallus gall
35	74.5	7.0	564	1 BMRE_CANAL	P28873 candida alb
36	74.5	7.0	1411	1 Y297_HUMAN	O15040 homo sapien
37	74.5	7.0	2145	1 U520_CAEEL	O90290 caenorhabdi
38	74	7.0	204	1 YAG0_ECOLI	P77262 escherichia
39	74	7.0	351	1 OPSD_SARDI	P79898 sarcocentro
40	73.5	6.9	153	1 YJEM_SALTY	Q92311 salmonella
41	73.5	6.9	276	1 NOSY_PSEST	P19845 pseudomonas
42	73.5	6.9	317	1 Y302_MYCGE	P47544 mycoplasma
43	73.5	6.9	461	1 YIFK_ECOLI	P27837 escherichia
44	73	6.9	362	1 CKRA_MOUSE	O91121 mus musculu
45	73	6.9	459	1 AAPA_BACSU	O06005 bacillus su

ALIGNMENTS

```

RESULT 1
RHTB_ECOLI
ID RHTB_ECOLI STANDARD; PRT; 206 AA.
AC P27847;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMOSERINE/HOMOSERINE LACTONE EFFLUX PROTEIN.
OS RHTB.
GC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=92338234; Pubmed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JUN-1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93313167; Pubmed=10386596;
RA Zakateva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
RA "Lyshits V.A.;"
RT "The novel transmembrane Escherichia coli proteins involved in the
amino acid efflux.";
RL FEBS Lett. 452:228-232(1999).
RN [4]
RP FUNCTION: CONDUCTS THE EFFLUX OF HOMOSERINE AND HOMOSERINE
LACTONE.
RN [5]
RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
RN [6]
RP SIMILARITY: BELONGS TO THE RHT FAMILY.
RN [7]
RP CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
BE INTRODUCED IN POSITIONS 60 TO EXTEND THIS ORF.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: M87049; AAA67620.1; ALT_FRAME.
EMBL: AE000458; AAC76827.1; ALT_FRAME.
PIR: S30714; S30714.
Ecogene: EG11469; rhtB.
InterPro: IPR001123; -.
Pfam: PF01810; lyse; 1.
Transport: Transmembrane.
TRANSMEM 5 25 POTENTIAL.

```

FT TRANSMEM 45 65 POTENTIAL.
 FT TRANSMEM 68 88 POTENTIAL.
 FT TRANSMEM 117 137 POTENTIAL.
 FT TRANSMEM 148 168 POTENTIAL.
 FT TRANSMEM 182 202 POTENTIAL.
 SQ SEQUENCE 206 AA; 22427 MW; 1164FI7738509C8C CRC64;

Query Match 84.7%; Score 898.5; DB 1; Length 206;
 Best Local Similarity 89.9%; Pred. No. 3.2e-77;
 Matches 187; Conservative 3; Mismatches 13; Indels 5; Gaps 3;

QY 1 MLEWMEFAYLLTSIIITLSPGSAINTMTSLNHGYPAGVYCWASDRG---DSYCAW 57
 1 MLEWMEFAYLLTSIIITLSPGSAINTMTSLNHGYPAGVYCWASDRG---DSYCAW 57
 Db 1 MLEWMEFAYLLTSIIITLSPGSAINTMTSLNHGYPAGVYCWASDRG---DSYCAW 57
 QY 58 RGVGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 117
 58 RGVGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 117
 Db 59 VGLGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 118
 59 VGLGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 118
 QY 118 NLTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 177
 118 NLTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 177
 Db 119 NLTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 178
 119 NLTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 178
 QY 178 PKOMKALNKIFGSLFMVGLLASARHA 205
 178 PKOMKALNKIFGSLFMVGLLASARHA 205
 Db 179 PKOMKALNKIFGSLFMVGLLASARHA 206
 179 PKOMKALNKIFGSLFMVGLLASARHA 206

RESULT 2

YEAS_ECOLI

ID YEAS_ECOLI STANDARD; PRT; 212 AA.

AC P76249; 007971; 007969;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-JUL-1998 (Rel. 36, Last annotation update)

DE HYPOTHETICAL 23.2 KDA PROTEIN IN GAPA-RND INTERGENIC REGION.

GN YEAS.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=K12;

RX MEDLINE=97251358; PubMed=9097040;

RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,

RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,

RA Mizubuchi K., Mori H., Mori T., Motomura K., Nakada S., Nakamura Y.,

RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,

RA Sivasubramanian S., Tagami H., Takeda J., Takemoto K., Wada C.,

RA Yamamoto Y., Horinouchi T.;

RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome

corresponding to the 40.1-50.0 min region on the linkage map.";

RL DNA Res. 3:379-392(1996).

RN [1]

RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

RC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 or send an email to license@isb-sib.ch).

DR EMBL; AEC00274; AAC74868.1; -
 DR EMBL; D90823; BAA15593.1; -
 DR EMBL; D90824; BAA15602.1; -
 DR Ecogene; EG13505; Yeas.
 DR InterPro; IPR001123; -
 DR Pfam; PF01810; Lyse; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 12 32 POTENTIAL.
 FT TRANSMEM 49 69 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 122 142 POTENTIAL.
 FT TRANSMEM 153 173 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 SQ SEQUENCE 212 AA; 23200 MW; 342B0D348C9AD9A CRC64;

Query Match 18.4%; Score 195.5; DB 1; Length 212;
 Best Local Similarity 28.6%; Pred. No. 2.1e-11;
 Matches 61; Conservative 45; Mismatches 82; Indels 25; Gaps 10;

QY 3 LEWMEFAYLLTSIIITLSPGSAI---NMTTSLNHGYPAG-GVYCWASDRGDSYC--A 55
 8 LNWMEFAYLLTSIIITLSPGSAI---NMTTSLNHGYPAG-GVYCWASDRGDSYC--A 55
 Db 8 LNWMEFAYLLTSIIITLSPGSAI---NMTTSLNHGYPAG-GVYCWASDRGDSYC--A 55
 QY 56 GWRGVGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 110
 56 GWRGVGTFSRSVIAFEVLKMGAAVLIWLGIOQWRAGAIDLKSLASTQSRRLFORAVEV 110
 Db 61 AMAGVATFLTKTPILFNIVYLCAGFYLLYGSKILVAT---LKGNSSEKSPDEPQYAI 116
 61 AMAGVATFLTKTPILFNIVYLCAGFYLLYGSKILVAT---LKGNSSEKSPDEPQYAI 116
 QY 111 FORAVEVINTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 169
 111 FORAVEVINTPKSIYFLALPFOEIMPPOPMOYIVLGVTTIVVDIIMIGVATLAORIALMIRK 169
 Db 117 EKRALILSLTPKPAILFYVSFEVQFDVNAHPHGISFLLAATLEVSFCYLSFLIISGA 176
 117 EKRALILSLTPKPAILFYVSFEVQFDVNAHPHGISFLLAATLEVSFCYLSFLIISGA 176
 QY 170 RLALMTKCPKPM-KALNKIFGSLFMVGLLAS 201
 170 RLALMTKCPKPM-KALNKIFGSLFMVGLLAS 201
 Db 177 FVQYIRTKKRLAKAGNSLIGLMEVGFARLAT 209
 177 FVQYIRTKKRLAKAGNSLIGLMEVGFARLAT 209

RESULT 3

YRHP_BACSU

ID YRHP_BACSU STANDARD; PRT; 210 AA.

AC 005406;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE HYPOTHETICAL 23.4 KDA PROTEIN IN AAPA-SIV INTERGENIC REGION.

GN YRHP.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;

OC Bacillus/Staphylococcus group; Bacillus.

OX NCBI_TaxID=1423;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=168;

RX MEDLINE=97453479; PubMed=9308178;

RA Kuster F., Ogasawara N., Yoshikawa H., Danchin A.;

RA Duesterhoeft A., Ehrlich S.D.;

RT "Sequence of the Bacillus subtilis genome region in the vicinity of

the lev operon reveals two new extracytoplasmic function RNA

polymerase sigma factors Sigv and Sigz.";

RL Microbiology 143:2939-2943(1997)

RN [2]

RP SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).

RC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z99117; CAB14652.1; -
DR EMBL: U93874; AAB80873.1; -
DR Subtilist; BG12304; yrhP.
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
SQ SEQUENCE 210 AA; 23389 MW; 2A37D9419FDB0A58 CRC64;

Query Match 18.4%; Score 195; DB 1; Length 210;
Best Local Similarity 27.6%; Pred. No. 2.3e-11;
Matches 56; Conservative 39; Mismatches 88; Indels 20; Gaps 6;

QY 8 AYLTSTIIITLSPGSAINTMTSLNHGYPAGVCMASDRFGDSCAGMR-----GVC 62
DB 6 AYPPIAMWVITPGADTLMVKNKTLRTG-PKAGRINILGATGTSF---WYVIALIGTSV 61
QY 63 LFSRSYIAFEVLKMGAAVILMLGIDQMRAGAIIDL-----KSLASTOSR-RHLFOR 113
DB 62 VIASKVILFTTITKYLGAALVILYGVSKFPFAKSMFLDDMQSAKMASSPKRYTSTMQ 121
QY 114 AVEFVNTLNKRSIVLEALFPQFIMPQPOLMOYIVLGVTIVDIIVMIGVTLAQRAL 173
DB 122 GSLSNILNKRKTLYVYVITMPOFINLNGINIOOLILILSLAVLWFLVYIIVYAKK 181
QY 174 WIKGPKOMALKNIFGSLFMWG 196
DB 182 WMKNSKFKQFKYKRTG--IILVG 202

RESULT 4
YB7_PSEAE STANDARD; PRT; 216 AA.
ID YB7_PSEAE
AC P38102; Q9H4V5; (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DE HYPOTHEICAL PROTEIN PA4757.
GN PA4757.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01.
RX MEDLINE=94222830; PubMed=8169201;
RA Kwon D.-H., Lu C.-D., Walthall D.A., Brown T.M., Houghton J.E.,
RA Abdelal A.T.;
RT "Structure and regulation of the carB operon in Pseudomonas
RT aeruginosa and Pseudomonas stutzeri: no untranslated region exists."
RL J. Bacteriol. 176:2532-2542(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U04992; AAA19047.1; -
DR EMBL: U81259; AAB39251.1; -
DR EMBL: AE004889; AAG08143.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 109 109 POTENTIAL.
FT CONFLICT 119 109 A -> G (IN REF. 1).
FT CONFLICT 119 119 P -> R (IN REF. 1).
FT CONFLICT 122 122 K -> Q (IN REF. 1).
SQ SEQUENCE 216 AA; 23248 MW; A30A08E7145918BD CRC64;

Query Match 17.0%; Score 180.5; DB 1; Length 216;
Best Local Similarity 28.0%; Pred. No. 5.3e-10;
Matches 60; Conservative 41; Mismatches 90; Indels 23; Gaps 9;

QY 4 EWMFAVLTSTIIITLSPGSAINTMTSLNHGYPAG-----GVYCMASDRFGDS--YCAG 56
DB 8 DFW-TYVIGVYVITLIGPNSLFLVLAISAGRGVATGRACGV-----LGDVLMILS 60
QY 57 WRGVGTLSRSYIAFEVLKMGAAVILMLGIDQ---WRAAGAIIDKSLASTOSRRLL-F 111
DB 61 ALGVAASLKAEPMLFGLKYLGAALVFLYLGGMARGAMKRLNDEARPAQAEQVDVQPF 120
QY 112 QRAVFNLTNKRISIVLEALFPQFIMP--QQPOLMOYIVLGVTIVDIIVMIGVATLAQ 169
DB 121 RKALLSLSNKRALIFETSPFQVDEYAVGL-STIVLVILIEDVSLATSLIFITGV 179
QY 170 RIALWIKGPKOMA-LNKIFGSLFMVGLDASA 202
DB 180 RLAMFRRRQRLAAGATSGVGLVGEVGLATA 213

RESULT 5
YG27_SYNY3 STANDARD; PRT; 206 AA.
ID YG27_SYNY3
AC P74343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 22.0 KDA PROTEIN SLR1627.
GN SLR1627.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneke T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirotsawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
RA Okumura S., Shimpō S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium

RA Nashimoto H., Saito N.:
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.,
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
 RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
 RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.,
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP IDENTIFICATION.
 RA Rudd K.E.;
 RL Unpublished observations (AUG-1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D13169; -; NOT_ANNOTATED_CDS.
 DR EMBL: D64044; -; NOT_ANNOTATED_CDS.
 DR EMBL: AE000344; AAC75631.1; -;
 DR EMBL: D90886; BA016464.1; -;
 DR ECoGene: EG12445; yf1K.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 8 28
 FT TRANSMEM 47 67 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT SEQUENCE 195 AA; 21248 MW; 5F86B828DDEC090 CRC64;
 SO

Query Match 11.9%; Score 126; DB 1; Length 195;
 Best Local Similarity 24.5%; Pred. No. 6e-05;
 Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;

QY 8 AYLLTSLITLTPSGAIIIMTSLNMGYPAGGVYCCASDRGDSVCACAGR----- 58
 DB 8 AFWTYITLITANTPEPNILALSSATSHGF-----RQSTRVLAISGLFLIVMLL 56
 QY 59 GVGTLFERSVY--AFEVLMWAGAAVYLWLGIQWRAGAIDKSLASTQSRHLLFORAV 115
 DB 57 CAGISFSLAVIDPAVAHLLSMAGAAIVWLAA--WKATISPTKED--GLQAKRISWASF 111
 QY 116 FVNLTNFKSVLFLAALFPQFIMPQQLMOMYIVLGVTTIVVIMIGVATLAQRIALMT 175
 DB 112 ALQFVNKKIILYGTALSTFVLP-QTQALSMVY-GVSV---LLAMIG--TFGN--VCNA 161
 QY 176 KGPQKMAKLNKINSLEMLYGALL 199
 DB 162 LAGHLFQRLFRQYGRQINIVLALL 185

RESULT 8
 RHTC_ECOLI

ID RHTC_ECOLI STANDARD: PRF: 206 AA.
 AC P27846:
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THREONINE EFFLUX PROTEIN.
 GN RHTC.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=92358234; PubMed=1379743;
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
 RT "Analysis of the *Escherichia coli* genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN [2]
 RP REVISIONS.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE OF 1-107 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=87115164; PubMed=3027506;
 RA Iritho N., Nakayama K., Nakayama H.;
 RT "The recQ gene of *Escherichia coli* K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [4]
 RP CONCEPTUAL TRANSLATION.
 RA Rudd K.E.;
 RL Unpublished observations (DEC-1997).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=99313167; PubMed=10386596;
 RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
 RA Livshits V.A.;
 RT "The novel transmembrane *Escherichia coli* proteins involved in the
 RT amino acid efflux.";
 RL FEBS Lett. 452:228-232(1999).
 CC -1- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
 CC INTRODUCED IN POSITION 80 TO PRODUCE THIS ORF.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
 CC IN POSITIONS 47 AND 73.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M87049; AA67619.1; ALT_FRAME.
 DR EMBL: AE000456; AAC76826.1; ALT_FRAME.
 DR EMBL: M30198; -; NOT_ANNOTATED_CDS.
 DR PIR: S30713; S30713.
 DR ECoGene: EG11468; rhtC.
 DR InterPro: IPR001123; -;
 DR Pfam: PF01810; Lyse; 1.
 KW Transport; Transmembrane.

DR PROSITE: PS00218; AMINO-ACID-PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 322 352 POTENTIAL.
FT TRANSMEM 356 400 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
SQ SEQUENCE 457 AA; 50215 MW; 5F2819526EA06BD8 CRC64;

Query Match 8.9%; Score 94; DB 1; Length 457;
Best Local Similarity 18.1%; Pred. No. 0.15;
Matches 50; Conservative 46; Mismatches 88; Indels 92; Gaps 10;

QY 3 LEWMPAYL-LTSIIITLSPSGAINMTTSLNHYAGGYCWASDRTGDSYCAGMRGV 61
DB 148 LEWMPFEPKATITIMIVAGFGII--IWGIGNGOGPTGIHNLMSN---GGFFSNGW--LG 200
QY 62 TLFRSRVIAFE-----VLKMGAAVLIWLGIOQ 89
DB 201 MWMSLQMWMPFAGGIEIGITAGEAKDPEKIPRAINSVPMRLVYVGTLFVI-MSIYP 259
QY 90 WRAAGAIIDKSLASTOSRRHLFORAV--FVNLTNPKSIV----- 126
DB 260 WNVGVGASPEVLTFOHMGITFAASILNFVLTASLSAINSDVFGVGRMLHGMAEGGSAP 319
QY 127 -----FLAALFPOFIMPQOPOLMQYVIVGTTIVDIIMIGY 164
DB 320 KVFAKTSRRGIPWTVLVMTTALLFAVLYINYPENVFIVIASLATFAFVWVMIMILSQ 379
QY 165 ATLQRIALMIKGPROMKALN-KIFGSLFMLVGALL 199
DB 380 IAFRRRLP-----PEVKALKFKVPGGVATTIGLI 410

RESULT 13
PROY_SALTY STANDARD; PRT; 456 AA.
AC P37460;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLINE-SPECIFIC PERMEASE PROY.
GN
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-LT2;
RX MEDLINE-97453475; PubMed-9308174;
RA Liao M.K., Gort S., Maloy S.;
RT "A cryptic proline permease in Salmonella typhimurium";
RL Microbiology 143:2903-2911(1997).
CC -!- FUNCTION: PERMEASE THAT IS INVOLVED IN THE TRANSPORT ACROSS THE
CYTOPLASMIC MEMBRANE OF PROLINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X74420; CA452441.1; -
DR PIR: S35983; S35983.
DR StyGene; SG10314; PROY.
DR InterPro; IPR002027; -.
DR Pfam; PF00324; aa-permeases; 1.
DR PROSITE: PS00218; AMINO-ACID-PERMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 332 352 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
SQ SEQUENCE 456 AA; 50047 MW; EED5485A676F7783 CRC64;

Query Match 8.7%; Score 92; DB 1; Length 456;
Best Local Similarity 18.5%; Pred. No. 0.23;
Matches 51; Conservative 46; Mismatches 87; Indels 92; Gaps 11;

QY 3 LEWMPAYL-LTSIIITLSPSGAINMTTSLNHYAGGYCWASDRTGDSYCAGMRGV 61
DB 148 LEWMPFEPKATITIMIVAGFGII--VWGIGNGOGPTGIHNLMSN---GGFFSNGW--LG 200
QY 62 TLFRSRVIAFE-----VLKMGAAVLIWLGIOQ 89
DB 201 MWMSLQMWMPFAGGIEIGITAGEAKDPEKIPRAINSVPMRLVYVGTLFVI-MSIYP 259
QY 90 WRAAGAIIDKSLASTOSRRHLFORAV--FVNLTNPKSIV-----NP 122
DB 260 WNVGVGASPEVLTFOHMGITFAASILNFVLTASLSAINSDVFGVGRMLHGMAEGGSAP 319
QY 123 K-----SIVFLAALFPOFIMPQOPOLMQYVIVGTTIVDIIMIGY 164
DB 320 KVFAKTSRRGIPWTVLVMTTALLFAVLYINYPENVFIVIASLATFAFVWVMIMILSQ 379
QY 165 ATLQRIALMIKGPROMKALN-KIFGSLFMLVGALL 199
DB 380 IAFRRRLP-----PEVKALKFKVPGGVATTIGLI 410

RESULT 14
LYSE_CORGL STANDARD; PRT; 236 AA.
AC P94633;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LYSINE EXPORTER PROTEIN.
GN
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-R127;
RX MEDLINE-97126810; PubMed-8971704;
RA Vrljic M., Sahm H., Eggeling L.;
RT "A new type of transporter with a new type of cellular function: L-
lysine export from Corynebacterium glutamicum";
RL Mol. Microbiol. 22:815-826(1996).
CC -!- FUNCTION: INVOLVED IN THE EFFLUX OF EXCESS OF L-LYSINE. THIS IS

NECESSARY TO CONTROL THE INTRACELLULAR L-LYSINE LEVEL.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN, INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE LYS/YESA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X96471; CAA65324.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; LysE; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
SQ SEQUENCE 236 AA; 25425 MW; EE86E8CF037C4727 CRC64;

Query Match 8.4%; Score 89.5; DB 1; Length 236;
Best local similarity 20.6%; Pred. No. 0.19; Mismatches 79; Indels 35; Gaps 4;
Matches 36; Conservative 25;

QY 3 LEWMAFALLSIILTSPGSAINTMTTSL-NHGYAGGVYCWASDRDSDSCYACGRGVG 61
DB 4 MEITFTGLIGASLLISGQNVLYIKQGIKREGLAIVLVCLISYF--LFIAGLQVD 61
QY 62 TLESRVIAEYVWKAGAYLWLGIOQWRAGADLKSLASTOSRRHLFQRAVEYNLTN 121
DB 62 LLSNAAYIVDIRMWGIAYLWFAVMAAKDA-----MTN 96
QY 122 PKSIFFALFPOFIMPOPOLMQYIVLGVTTIVDIIYMGVATLQRIALMIK 176
DB 97 KVERA-----PQIIEETPTVPDDTPLGSAVADTRNRREVSVDKORVWK 144

RESULT 15
YD07_HAEIN
ID YD07_HAEIN STANDARD; PRT; 210 AA.
AC 057320; 005057;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN H11307.
GN H11307.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=9550630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shurely R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Rine L.D., Fritchman J.L., Fuhman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32810; AAC22954.1; -
DR TIGR: H11307; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; LysE; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 66 86 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
SQ SEQUENCE 210 AA; 23636 MW; 254D159014845473 CRC64;

Query Match 8.2%; Score 87; DB 1; Length 210;
Best local similarity 21.1%; Pred. No. 0.29; Mismatches 66; Indels 18; Gaps 5;
Matches 32; Conservative 36;

QY 59 GVGTLFSRSVIAFEYWKAGAYLWLGIOQWRAGADLKSLASTOSRRHL-----FOR 113
DB 56 GLAVLFVTPALHGYIMLGSGYLAYLGFIMARSKKAKFESHSDTEHQOTTIKKELK 115
QY 114 AVFVNLTPKSVIFLALFPQFIMPOPOLMQYIVLGVTTIVD-----ITVIGATL 167
DB 116 GLVNLISNAKVVYFSSVW-SLVVNITEEMQ--ITLFAVIVETFCFYVYISLFSRNI 173
QY 168 AQRALMIKSPKQMKALKIKFGSLFMLGALL 199
DB 174 AKRLY-----SQSKYIDNMAGIYFLFFGCVL 200

Search completed: May 6, 2001, 14:41:20
Job time: 191 sec

```
Db 116 IFVNLTPKSTVFLEALFPQFIMPOQPOLAOYLITGVTTIVDMVYNTGYATLQRIAN 175
Qy 175 IKGPKOKALKINKIFGSLFMLEVGCALLASAKRA 205
Db 176 IKGPKOKALKINKAFGSLFMLEVGCALLASAKRA 206

RESULT 2
Q9KVK7 PRELIMINARY; PRT; 205 AA.
AC Q9KVK7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE HYPOTHEITICAL PROTEIN VCO136.
GN VCO136
OS Vibrrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrrio.
OX NCBI_TaxID=666;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Peterson J.D., Mayam L.A.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Olin H., Dragoi J., Sellers P.,
RA McDonald L., Ueberback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AE004104; AAF93131.1; --
DR TIGR; VCO136; --
SQ SEQUENCE 205 AA; 22099 MW; AE42CA6FA6EA1223 CRC64;

Query Match 43.3%; Score 459.5; DB 2; Length 205;
Best Local Similarity 43.4%; Pred. No. 3.3e-33;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;

Qy 1 MTLFMPFAYLTSTIIITLSPGSAINTMTSLNHG--YPAGVGYCWASDRGSDYCAQWR 58
Db 1 MDIHWMLAYLTAVFSLAPSGIVNSISNGLSYGFHSHSGAIIIGIQLACHIVIG-I 59
Qy 59 GVGTLFSRVIAFEVYKMGAAVYLTNGIQOMRAAGADLKSASQSRHLPQRAFEVN 118
Db 60 GIGALVQSLATFTLIKMGAAVYLTNGIQOMRAAFUTTTSHLSQAALRAVAVIN 119
Qy 119 LTNKSTVFLEALFPQFIMPOQPOLMOYIVLGTTIVVDIIVMGVATLQRIALMKIP 178
Db 120 LTNKSTVFLEALFPQFIMPOQPOLMOYIVLGTTIVVDIIVMGVATLQRIALMKIP 179
Qy 179 KOKALKINKIFGSLFMLEVGCALLASAR 203
Db 180 NIMTRMKLFGSMFGCGMLATATK 204

RESULT 3
Q9RSX2 PRELIMINARY; PRT; 241 AA.
AC Q9RSX2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CONSERVED HYPOTHEITICAL PROTEIN.
GN DR1999.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN 11
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=R1;
RA MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwin M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Olin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Ueberback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioreistant Bacterium Deinococcus
RT radiodurans R1";
RL Science 286:1571-1577(1999).
DR EMBL; AE002037; AAL1548.1; --
DR TIGR; DR1999; --
DR INTERPRO; IPR000719; --
DR PFAM; PF01810; Lyse; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN.1.
SQ SEQUENCE 241 AA; 25161 MW; A4017ABFD860374 CRC64;

Query Match 21.5%; Score 228; DB 2; Length 241;
Best Local Similarity 30.3%; Pred. No. 1.1e-12;
Matches 64; Conservative 41; Mismatches 84; Indels 22; Gaps 6;

Qy 8 AVLTSTIITLSPGSAINTMTSLNHGYPAGVGYCWASDRGSDYCAQWR-----G 59
Db 38 AFLVAAYVALLPQGLMYTLASLOGGR-----WAGLSALGTAGAGWVHLASAVG 90
Qy 60 VGTLSRVIAFEVYKMGAAVYLTNGIQOMRAAGADLKSASQSRHLPQRAFEVN 114
Db 91 ISALIMASSLAFSVYKMGAAVYLTNGIQOMRAAGADLKSASQSRHLPQRAFEVN 150
Qy 115 VEVNLTNPKSTVFLEALFPQFIMPOQPOLMOYIVLGTTIVVDIIVMGVATLQRIAL 173
Db 151 AMTELINPKSTVFLEALFPQFIMPOQPOLMOYIVLGTTIVVDIIVMGVATLQRIAL 210
Qy 174 WIKG-PKOKALKINKIFGSLFMLEVGCALLASAR 203
Db 211 RLGQNPBRROGOKVASCAGMALTGVAVER 241

RESULT 4
Q9R6F8 PRELIMINARY; PRT; 204 AA.
AC Q9R6F8;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE TIORF101 PROTEIN.
GN TIORF101.
OS Agrobacterium tumefaciens (Ti plasmid pTi1737).
OC Plasmid pTi-SAKURA.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
OX NCBI_TaxID=362;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";
RL Gene 242:331-336(2000).
RN 12
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopline-type trbu
RT gene: construction of genetic and physical map and sequencing of
RL trb/trai and rep gene clusters of a new Ti plasmid pTi-SAKURA";
RN 13
RP SEQUENCE FROM N.A.
```

RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid."; Ser. 37:159-160(1998).
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of
RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.";
RL Nucleic Acids Symp. Ser. 39:265-266(1998).
DR EMBL: AB016260; BAA87726.1; -
DR INTERPRO: IPR001123; -
DR PFAM: PF01810; Lyse; 1.
KW Plasmid
SQ SEQUENCE 204 AA; 21015 MW; A6E2AE267E0E97D0 CRC64;

Query Match 18.4%; Score 195.5; DB 2; Length 204;
Best Local Similarity 24.4%; Pred. No. 6.7e-10;
Matches 55; Conservative 44; Mismatches 83; Indels 43; Gaps 6;
QY 1 MTEWMAVLLTSLITLSPGSGAINTMTTSLNHGYPAGVYCWASDRITGDSYCAQWRCV 60
DB 1 MDISTLAFAPAAFEVFAASPGDPNMTIVARTISHGAASGATV-----GA 44
QY 61 GTL-----FNSVIAFE-----VLKAGAAVLIWIGIQWRAGAT-DIKSLA 102
DB 45 GTVAGILLFTTLAFAFGSLTAEMGIWMTLRYGAAVLIWIGIKMTAEPPVPELOPVS 104
QY 103 STGSRRLFORAVENVLTNPKSIVFLAALPQPI---MPQOQLMQYIVLGVTTIVDI 158
DB 105 ERRLGLAVFAIGVALNIGNKMPLEFYALLPNVVGSSLPFGHTAIAAAYIL-----AVEM 159
QY 159 IVMIGYATTLAQRIALWIKGPKOMKALNKIFGSLFMLVGALLASAR 203
DB 160 VVIGGHVILAGRARAKLLRPTTVVRVNRRTAGVGAVVAVNASR 204

RESULT 5
Q9KFP9 PRELIMINARY; PRT; 207 AA.
AC Q9KFP9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE BH0429 PROTEIN.
GN BH0429.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001508; BAB04148.1; -
SQ SEQUENCE 207 AA; 22317 MW; 9FED148C8E8DD3F0 CRC64;

Query Match 18.1%; Score 192.5; DB 2; Length 207;
Best Local Similarity 26.3%; Pred. No. 1.2e-09;
Matches 57; Conservative 50; Mismatches 87; Indels 23; Gaps 6;
QY 1 MTEWMAVLLTSLITLSPGSGAINTMTTSLNHGYPAGVYCWASDRITGDSYCAQWRCV 58
DB 1 MDISTLSPGSAVALLTTPGPDILFVLAQSMQKQAGIV-----TALGLCTGLLVH 53
QY 59 -----GVGTFEFSRVIAFEVLKWAAGVLIWIGIQWRAG--AIDLKSLASTQSRRH 109
DB 54 ITAAVVGISALIVQSALAFYVYKGAAYLLYLAWKAFQEKGEGLSDKQ---TTLAVGA 110
QY 110 LQGRAVENVLTNPKSIVFLAALFPQFIMPQOPQ-LMQYIVGVYTTIVDIWIGYATLA 168
DB 111 LYKKGITNMVNLPKVSLFLLALPQFVNSGAGSAPWOMLLLGAVFLIQAFITFSLVSWFA 170
QY 169 ORIA-LWIKGPKOMKALNKIFGSLFMLVGALLASARH 204
DB 171 EVVQGLMRSSFIKQMRIRKGGLLALIGLYVAFSKN 207

RESULT 6
Q9R6J5 PRELIMINARY; PRT; 229 AA.
AC Q9R6J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE TIORF64 PROTEIN.
GN TIORF64.
OS Agrobacterium tumefaciens (Ti plasmid pTiT37).
OC Plasmid pTi-SAKURA.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
OX NCBI_TaxID=362;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";
RL Gene 242:331-336(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopline- type trbu
RT gene: construction of genetic and physical map and sequencing of
RT trb/trai and rep gene clusters of a new Ti plasmid pTi-SAKURA.";
RL Biochim. Biophys. Acta 1396:1-7(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid.";
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;

OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
 OC Zymomonas;
 OX NCBI_taxid=542;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=ZM4;
 RA Um H.W., Kang H.S.;
 RT "The sequence analysis of 42D7 fosmid clone of Zymomonas mobilis
 RT ZM4.";
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF157493; AD42407.1; -
 DR INTERPRO: IPR001123; -
 DR PFAM: PF01810; LysR; 1
 KW Hypothetical protein.
 SO SEQUENCE 158 AA; 17454 MW; FE44A9639DA75A64 CRC64;

Query Match:	17.4%;	Score	184.5;	DB	2;	Length	158;
Best Local Similarity	30.1%;	Pred.	No. 4,76-09;				
Matches	46;	Conservative	34;	Mismatches	58;	Indels	15; Gaps
OY	59	GVGTLFSRSVIAFEVKMGAGAYLLIMLGQ-----QWTAAGADILKSLASTOSRHFLFOR	113				
Dd	11	GLCAVLRASTALANLTKMSGAAIYLCLWLGTINLIHPRKLVNDLDPSMPSATKALRGCF--	68				
OY	114	AVFVNLTNPKSIVLFELAAPPOFIMPQQEQOLMQOYVLGVTITVDDIYMIQAVTIAORIA-	172				
Dd	69	--FTNLINLPVGVIEFYVTLPGF-MRSSHLIQYNFL-LFFIHMTITIIIMFALLATATAP	123				
OY	173	--LMIGPKOMAKLNKIIFSGLSFMLVGAALLAASR	203				
Dd	124	LIRFELKRPRELVSIDRLTGCVFIARGIKLATASK	156				

RESULT	10
Q9RMX0	
ID	Q9RMX0
AC	Q9RMX0;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE	PMO2-63.
OS	Bacillus anthracis.
OG	Plasmid PMO2.
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;
OC	Bacillus/Staphylococcus group; Bacillus.
OX	NCBI_Taxid=1392;
ON	[1]
RP	SEQUENCE FROM N.A.
RA	Okina S.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA	Kumano R.T., Menter D., Martinez Y., Svensson R., Tatum L.R.,
RL	Brown A.E., Jackson P.J.;
DR	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL: A0188935; AAF13668.1; -
DR	INTERPRO: IPR001123; -
DR	PFAM: PF01810; Lyse; 1.
KW	Plasmid.
Q0	SEQUENCE. 205 AA; 22900 MW; 6DC91EE968B524B8 CRC64;

	Query Match	17.3%	Score 184	DB 2	Length 205
	Best Local Similarity	29.5%	Pred. No. 7e-09		
	Matches 64	Conservative 36	Mismatches 83	Indels 34	Gaps 10
QY	1 MLEWMEFAVLLLSILTLSPGSGAINTMTTSLNIGYPAG--GYCCMASDRGD-----SYCA 55				
Db	1 MDINTITVLLVLTGTVSPGPNMAVITIKNSL---YRSLSGVSTVAGIAGSLIHITVYCL 57				
QY	56 GMRGVSTLPSRSVIAFEYLKRWAGAAYLLWLTSLIQOMRA---AGALDKSLASTOSRRHQL 112				
Db	58 --IGIVIIISKSLLEFNLTKWIGVAYLYLTGIKLLRSKRSKSPAAIIKNNESTWK--AFR 113				
QY	113 RAEVVALTNPKSIVELALPQFIIMPQDPQLMOYIVLGVTIVVDII---VMIGATTLAQ 169				

```
Db      114 SEFTLDMNPKATLEYLAIFQVIEPNTNIIVQ-SYGLTWASVEIIIMHVLVEFETHKS 172
        ::|||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Qy      170 -----RIALNIKGPQKMAKNKEGSLFMWGAALLA 200
        :|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db      173 VRNFYSTISHMTE-----RYTGALLILGLIRLA 200
```

RESULT	11
Q9P9Z0	
ID	Q9P9Z0
AC	Q9P9Z0;
DT	01-OCT-2000 (TREMBLrel. 15, Created)
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT	01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE	AMINO ACID TRANSFERER.
CN	XF2730.
OS	Xylella fastidiosa.
OC	Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group
CC	Xylella.

DR EMBL; AE004079; AAF85515.1; -.
DR INTERPRO; IPR001123; -.
PEAM: PF01810; IysE; 1.
DR
SEQUENCE 213 AA: 23614 MW; E97242DD507C8065 CRC64;

[illegible]

AC O9KLA0:
DT 01-OCT-2000 (TREMblRel. 15, Created)
DT 01-OCT-2000 (TREMblRel. 15, Last sequence update)
DT 01-OCT-2000 (TREMblRel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN VCA0846.
GN VCA0846.
OS *Vibrio cholerae*.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; *Vibrio*.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; EpubMed=10932301;
RA HeideIdberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragol I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith N.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen *Vibrio*
RT *cholerae*.";
RL Nature 406:477-483(2000).
DR EMBL: AE004412; AAF6744.1; -.
TI GR: VCA0846; -.
SQ SEQUENCE 204 AA; 21719 MW; D09887299659FBD CRC64;

1

2

3

alignment_block:
US-09-466-935-2 x AQ859250/rev ..

Align seg 1/1 to reverse of: AQ859250 from: 1 to: 473

```
1 MetThrLeuGluTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuTh 17
   |||||
229 ATGACCTTAGATGTTGGTTGCTACCTGCTGACATGCATCATTTTAAAG 180
   |||||
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
   |||||
179 CCTGTGCACAGCTCTGCTGCATCAACACATCATACACCTCGCTCAAC 130
   |||||
34 IsGlyTyrProAlaGly..GlyValTyrCysTrpAlaSerAspArgThrG 50
   |||||
129 ACGGTTATCGCGGCGCATAAAGTCTATGCTGGGCTTCAGACCGGACTG 80
   |||||
50 LysAspSer 52
   |||||
79 GCGATTTCG 72
```

seq_name: gb_est95:BG038404

seq_documentation_block: 443 bp mRNA EST 24-JAN-2001
LOCUS BG038404
DEFINITION dg34h08.y1 Xenopus laevis gastrula non normalized Xenopus laevis
CNA clone XENOPUS.SOURCE.ID:xlnnga009p16 5' similar to TR:Q9RSX2

ACCESSION BG038404
VERSION BG038404.1 GI:12480989

KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 443)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Other ESTs: dg34h08.x1
Contact: Sandy Clifton, Ph.D
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

LIBRARY constructed by Bruce Blumberg
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> Please reference the id listed
below when ordering this clone: Source lab clone id - xlnnga009p16
Seq primer: -40RP from Gldcoo
High quality sequence stop: 430.
Location/Qualifiers

FEATURES
Source
1..443
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS-SOURCE.ID:xlnnga009p16"
/clone_lib="Xenopus laevis gastrula non normalized"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae)
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library construction by Bruce Blumberg (Cho et al. 1991 Cell 67, 1111-1120). Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 74 a 142 c 127 g 100 t
ORIGIN

alignment_scores:
Quality: 180.50 Length: 141
Ratio: 1.920 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 28.369

alignment_block:
US-09-466-935-2 x BG038404 ..

Align seg 1/1 to: BG038404 from: 1 to: 443

```
9 TyrLeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaI 25
   ::|||
14 TTCCTAATACATCTTCATATGTTGCCCGCTCCGGCAGCGGAGATG 63
   |||||
25 easnThrMetThrThrSerLeuAsnHsGlyTyrProAlaGlyValT 42
   ::|||
64 TTACACGCTGCTGCCGCGCTTTCGACAGGGGCGCGGACGTCATCG 113
   |||||
42 yrcysTrpAlaSerAspArgThrGlyAspSerTyrCys..... 54
   |||
114 CGGCTTTCGGC.....TGCACGCTGGGCATC 139
   |||||
55 .....AlaGlyTyrArgGlyValGlyThrLeuPheSerAr 66
   |||||
140 GNGCCGATCTGCTCGCGCCATCACCGGCTTTCGCCCATCTGCACAC 189
   |||||
66 gserValIleAlaPheGluValLeuLysTrpAlaGlyAlaTyrLeu 83
   |||||
190 CAGTGGCGGCTTTCAGCGTGTCAATATCTGGGGGCGCTATCTGC 239
   |||||
83 lertPLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspleuLys 99
   |||||
240 TCTACATGGCGGTGAACACCGTCGACAGACATGCGCTGAAGATCAT 289
   |||||
100 SerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValP 116
   |||||
290 GAACGAGGCGACCGCCGAAAGCCGCCGTCATCGGTGAGCGCATCT 339
   |||||
116 eValAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPhe 133
   |||||
340 CATCAATATGCTGCAACCCGAAACTGTCATCTTTTGGCTTCCTGC 389
   |||||
133 roGlnPheIleMetProGlnGln 140
   |||||
390 CGCAGTTCAATCGCGCGCATGAG 412
```

seq_name: gb_est89:Bf614897

seq_documentation_block: 428 bp mRNA EST 14-DEC-2000
LOCUS Bf614897
DEFINITION dg18a02.y1 Xenopus laevis gastrula non normalized Xenopus laevis
cDNA clone XENOPUS.SOURCE.ID:xlnnga005d04 5' similar to TR:Q9RSX2
Q9RSX2 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA sequence.
ACCESSION Bf614897 GI:11787968
VERSION Bf614897.1
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodidae; Xenopus.

TITLE
JOURNAL
COMMENT

1 (bases 1 to 428)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu

Library constructed by Bruce Blumberg
Library normalized by: Washington University Genome Sequencing Center
DNA Sequencing by: Washington University
Clone distribution: Xenopus clone distribution information for this library can be found through Research Genetics, visit their web page at: <http://www.resgen.com/> Please reference the id listed below when ordering this clone: Source lab clone id - xlnhg005b04
Seq primer: -40bp from Gldco
High quality sequence stop: 401.

FEATURES
source

1..428
Location/Qualifiers
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS_SOURCE.ID:xlnhg005b04"
/clone_lib="Xenopus laevis gastrula non normalized"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae)
EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss phagemids were made double stranded and electroporated into Top-10 F. Original library construction by Bruce Blumberg (Cho et al. 1991 Cell 67, 111-1120). Note: This is a Xenopus Gene Collection (XGC) library."
BASE COUNT 67 a 137 c 127 g 96 t 1 others
ORIGIN

alignment_scores:

quality: 164.50 length: 135
Ratio: 1.848 Gaps: 2
Percent Similarity: 65.926 Percent Identity: 27.407

alignment block:

US-09-466-935-2 x BF614897 ..

Align seg 1/1 to: BF614897 from: 1 to: 428

15 ldeuThrleuSerProGlySerGlyAlaIleAsnThrMetThrTrp 31
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3 ATGCTTGCCGCGCTCTCGGCGACCGCGTGTACACGCGTGGCGG 52
31 rleuAsnHISGlyTyrProAlaGlyAlaValTyrCysTrpAlaSerAsp 48
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
53 CCTTTCAGAGGCGCGCGCGAGCGATGATGCGCTTCGCGC..... 95
48 rgrHrGlyAspSerTyrCys.....Ala 55
|||

FEATURES
source

96TGCAGCTCGGCATCGTGGCGCATCTGCTCGCC 128
56 GlyTrpArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheG1 72
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
129 GCATCATCGCGTCTTGCGCGCATCTGCACACAGAGCGCTGGCCTTNG 178
72 uValleuLysTrpAlaGlyAlaValTyrLeuIleTrpLeuGlyIleGing 89
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
179 CGTGTCAAAATATCTGGGGCTGATCTGCTCATCATGGCGTGACACA 228
89 lntPrpAlaAlaGlyAlaIleAspLeuLysSerLeuAlaSerTrpGin 105
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
229 CGCTGCAGAGAAATGCGCGCTCAAGATGCATAAACGAGCGACGCCCA 278
106 SerArgTrpHisLeuPheGlnArgAlaValPheValAsnLeuThrAsnPr 122
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
279 AAACCCCGCCGCGCATCGGTGAGCGGATCTCATCACTGCTGAAACC 328
122 oLysSerIleValPheLeuAlaIleAlaLeuPheProGlnPheIleMetProG 139
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
329 GAACGTGCGATCTCTTTTTCGCTTCCTCGCGCAGTTCATCGCGCGG 378
139 lngIn 140
:||||:||||:||||:||||:||||:||||:||||:||||:||||:
379 ATGAG 383
seq_name: gb_gss5:AQ365906
seq_documentation_block: 562 bp DNA GSS 16-DEC-1999
LOCUS AQ365906
DEFINITION nbx0064N05f CUGI Rice BAC Library Oryza sativa genomic clone
nbx0064N05f, DNA sequence.
ACCESSION AQ365906
VERSION AQ365906.2 GI:6585776
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
REFERENCE 1 (bases 1 to 562)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT On Dec 15, 1999 this sequence version replaced g1:4215558.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: twing@clemson.edu
Seq primer: TAATACGACTCACTATAGGG
Class: BAC ends
High quality sequence stop: 400.
Location/Qualifiers
1..562
/organism="Oryza sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbx0064N05f"
/clone_lib="CUGI Rice BAC Library"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBel0BAC11; Site.1: HindIII; Site.2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocultivorous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small

genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 91 a 151 c 180 g 140 t
ORIGIN

alignment_scores:
Quality: 149.50 Length: 108
Ratio: 2.534 Gaps: 2
Percent Similarity: 54.630 Percent Identity: 27.778

alignment_block:
US-09-466-935-2 x AQ365906 ..

Align seg 1/1 to: AQ365906 from: 1 to: 562

```

1 MethrLleuGluTrpTPheAlaTyrlleuThrSerlleleuTh 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
299 ATGTCGCGTGAACGCGTGGCTTTTCGCCCTGTTGGTATCAG 348
17 rleuSerProGlySerGlyAlaIleAsnThrMetThrSerleuAsn 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
349 TCTTCTCCGGCGCCAGCGCATTCGTCGATCCAGTGCAGT 398
34 ISGlyTyPrAlaGlyGlyValTyrcyStrPaAsaPaTyThGly 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
399 ACGGCTTC.....TGGCGCGTTACTGAAACGCCCTCGGTGCA 406
51 AspSerTyrcysAlaGlyTyPrAg..... 58
407 .....TGGCGCGTTACTGAAACGCCCTCGGTGCA 438
59 .....GlyValGlyThrL 63
439 ATTGGCGCTGATCGTTCAGATCGCATCAGCCGCGAGCTGGTCCG 488
63 eupheSerArgSerValIleAlaPheGlyValleuTyStrPaAlaGlyAla 79
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
489 TGCTGCGCCAGCTCGCGCTGCGCTTCAAGCTGATCAATGTTGCGGCTT 538
80 AlaTyrlleuIleTrpLleuGlyIle 87
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
539 GCTTACCTGGTGTATCTGGGTGTG 562

```

seq_name: gb_gss12:A0858095

seq_documentation_block:

LOCUS A0858095 769 bp DNA GSS 03-NOV-1999
DEFINITION nbe001114f CGI Rice BAC Library (EcoRI) *Oryza sativa* genomic
clone nbe001114f, DNA sequence.

ACCESSION A0858095
VERSION A0858095.1 GI:6208552

KEYWORDS GSS.

SOURCE *Oryza sativa*.

ORGANISM *Oryza sativa*

Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae;
Oryza.

REFERENCE 1 (bases 1 to 769)

AUTHORS Wing,R.A. and Dean,R.A.

TITLE A BAC End Sequencing Framework to Sequence the Rice Genome

JOURNAL Unpublished (1998)

COMMENT Contact: Wing RA

Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGCG
Class: BAC ends
High quality sequence start: 42
High quality sequence stop: 454.

FEATURES

source

1..769
/organism="Oryza sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe001114f"
/clone_lib="CGI Rice BAC Library (EcoRI)"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo, Site_1: EcoRI, Site_2: EcoRI.
Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety using EcoRI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 203 a 216 c 176 g 169 t 5 others
ORIGIN

alignment_scores:
Quality: 126.00 Length: 36
Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 91.667

alignment_block:
US-09-466-935-2 x A0858095 ..

Align seg 1/1 to: A0858095 from: 1 to: 769

```

1 MethrLleuGluTrpTPheAlaTyrlleuThrSerlleleuTh 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
658 ATGACCTTAGAATGTGGTT..GCTTACCTGTCGATCATGATCAT..TTAAG 705
17 rleuSerProGlySerGlyAlaIleAsnThrMetThrSerleuAsn 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
706 CCTGTCGCCAGGCTCTGTGTCATCAACACATATGACCACTCGCTACCC 755
34 ISGlyTyPr 36
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
756 ACGGCTAT 763

```

seq_name: gb_gss23:AZ576789

seq_documentation_block:

LOCUS AZ576789 407 bp DNA GSS 08-DEC-2000
DEFINITION 01a12 Shot-gun genomic library of *Rhizobium* strain ANU265 *Rhizobium*
sp. NGR234 genomic clone 01a12, DNA sequence.

ACCESSION AZ576789


```

304 TGG.....AGAAATGGCTGGAGCTTTTGT..... 327
60 lGlyThrLeuPheSerArgSerValIleAlaPhe.GluValLeuLysTrp 76
328 .GGCAGAGTGTCTTACTACCAATGTT.....CTATGAGGACTCAAG... 367
77 AlaGlyAlaAlaTyTrpLeuIleTrpLeuGlyIleGlnGlnTrpArgAla 93
367 ..... 367
93 aGlyAlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgArgHisL 110
368 .....ATAGCCCGAGAGAGCCTGCTCGTAAGTCACAAGTCAGGCATT 410
110 euPheGlnArgAlaValPhe.ValAsnLeuThrAsnProLysSerIleVa 126
411 GCCTACAA.....TTCATGCTGCTGCCAGGACCAAAATGGAACCAT 451
126 lPheLeuAlaAlaLeuPheProGlnPheIleMetProGlnGln..... 140
452 CTTTATGGAGACA.....CACAAACTGTTGGCAACAGATGCTGA 492
141 .....ProGlnLeuMetGlnTyIleValLeuGlyValThrThrIleVal 155
493 GCTTTCCTCAGCTCCTGAAA...ACAGTGCTGCACATCATCCAGGTGGHC 539
156 ValAspIleIleValMetIle..... 162
540 ATAAGCTACTTCTCATGCTCATCTTCATGAACACTACACGGGTACCTCG 589
163 .....GlyTyAlaThrLeu..... 167
590 CATTCAGTACAGCAGCAGCGGGGTACAGGATACATTACTCTTACGCTGG 639
168 .....AlaGlnArgIleAlaLeuTrpIleLys 176
640 ACAAGAGAGTGGTGTAGTGGGTATCACAGGGTTCGCTGTGATTA 689
177 GlyProLysGlnMetLysAlaLeuAsnLysIlePheGlySerLeu 191
690 GGCATAGGGGGGTGTTTCGAATGAGGGGAGATTTTGGACACTTG 734

seq_name: gb_est88:BF526220

seq_documentation_block:
LOCUS BF526220 774 bp mRNA EST 11-DEC-2000
DEFINITION 602071232F1 NCI_CGAP_Brn64 Homo sapiens cDNA clone IMAGE:4214263
5', mRNA sequence.
ACCESSION BF526220
VERSION BF526220.1 GI:11613570
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 774)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM9787 row: p column: 08
High quality sequence stop: 701.
Location/Qualifiers
FEATURES
```

```

source 1..774
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Brn64"
/tissue_type="glioblastoma with EGFR amplification"
/lab_host="DH108 (T1 phage-resistant)"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.57 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 148 a 198 c 231 g 196 t 1 others
ORIGIN

alignment_scores:
  Quality: 90.50 Length: 195
  Ratio: 1.065 Gaps: 8
  Percent Similarity: 43.590 Percent identity: 25.128

alignment_block:
US-09-466-935-2 x BF526220 ..

Align seg 1/1 to: BF526220 from: 1 to: 774
```

```

21 GlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHis..... 34
287 GCATGAGNACTTCAANTCAATGTCGCCACAGTCAGGGAGTCGGGA 336
35 .....GlyTyProAlaGlyGlyValTyCysTrpAlaSerA 47
337 GGTAGGGCTGTAGGCCAACCAATTGGTGGAGTCTCAGCGATAGCCAGG 386
47 sPArgThrGlyAspSerTyCysAlaGlyTrpArgGlyValGlyThrLeu 63
387 TGAGAAAGTGGTTCACCCAGGGGAGGGTGG.....GGCCTCGG 427
64 PheSerArgSerValIleAlaPheGluValLeuLys.....Tr 76
428 GCAGATCTGTCCTCTTGGCCCTCTGTCTCTCAATGTCCAAATGTGTG 476
76 palAGlyAlaAlaTyTrpLeuIleTrpLeuGly....IleGlnGlnTrpArgA 92
477 GAGGACCTCTGTTCATATCCCGCTGGCTCTTCCAGCAGTGGAGTT 526
92 laAlaGlyAlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgArg 108
527 ACTGTAGAGGGATG..... 540
109 HisLeuPheGlnArgAlaValPheValAsn.....LeuThrAs 121
541 .....TCCCAAGCTGTGTTTCCCAATCAGGTTAAGCTGTTGTGAA 581
121 nProLysSerIleValPheLeuAlaAla..... 130
582 CTCTCCGCTGTGTGTGNTGTTGGTGGTGTGTGAGACACATCAGTG 631
131 .....LeuPheProGlnPheIleMetProGlnGlnProGlnLeu 143
632 TGTGCAGGCTGTGTTCCCATTTCTCTCTCTCCCTTCCAGACCATCATTTG 681
144 MetGlnTyTrpIleValLeuGlyValThrThrIleValAlaAspIleIleVa 160
682 AGGAACAAATGTAAGAAATCCGTCCCAATC..... 714
160 lMetIleGlyTyAlaThrLeuAlaGlnArgIleAlaLeuTrpIleLysG 177
715 .....TCCTGGCTCCAG 727
177 lyProLysGlnMetLysAlaLeuAsnLysIlePhe 188
728 GGCCTCTCGGGGGAACAGATCACCAGGATCTTC 762
```

seq_name: gb_est53:AW980930

seq_documentation_block: 607 bp mRNA EST 07-SEP-2000
LOCUS AW980930
DEFINITION EST392083 GVN Medicago truncatula cDNA clone pGVN-60123, mRNA sequence.

ACCESSION AW980930
VERSION AW980930.1 GI:8172475
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids 1; Fabales; Fabaceae; Papilionoideae; Medicago.

REFERENCE
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S., Holt,I.E. and Fraser,C.M.
TITLE ESTs from one month old nitrogen-fixing root nodules of Medicago truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612-625-7219
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
TIGR sequence name: MTCBW60NK
More information is available at: . http://chrysis.tamu.edu/medicago
Seq primer: SKmod (CTA gAA CTA gTg gAT CC).

FEATURES
source
1..607
/cultivar="Medicago truncatula"
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="pgVN-60123"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stage="effective root nodules harvested one month post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XL0LR"
/notes="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from polyA+ enriched RNA from effective root nodules harvested one month post inoculation with Sinorhizobium meliloti. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells."

BASE COUNT 172 a 161 c 133 g 141 t
ORIGIN

alignment_scores:
Quality: 89.00 Length: 202
Ratio: 0.918 Gaps: 12
Percent Similarity: 48.020 Percent Identity: 22.772

alignment_block:
US-09-466-935-2 x AW980930/rev ..

Align seg 1/1 to reverse of: AW980930 from: 1 to: 607

48 ArgThrGlyAspSerTyrCysAlaGly.....Tr 57
||| |||:||||:||||: |||
569 AGGAGGGGTCGCTCATGTGGGGGGAGGGTATCCACCACTGCATTTG 520
|||:||||:||||:||||: |||
57 pArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluVal 74
|||:||||:||||:||||: |||
519 GAGGAGCATATCCACCTTGATTTGGAGGGCGTATCCGCC.....T 479

us-09-466-935-2.rst

Mon May 7 11:08:09 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 15:09:39 ; Search time 1239.41 Seconds
(without alignments)
7353.642 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804
Perfect score: 618
Sequence: 1 atgttgatgtatttcttcac.....ttcatttgattatttcggg 618

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba1.*
 - 2: gb_ba2.*
 - 3: gb_ba3.*
 - 4: gb_in1.*
 - 5: gb_in2.*
 - 6: gb_in3.*
 - 7: gb_om.*
 - 8: gb_ov.*
 - 9: gb_pat1.*
 - 10: gb_pat2.*
 - 11: gb_ph.*
 - 12: gb_pl1.*
 - 13: gb_pl2.*
 - 14: gb_pl3.*
 - 15: gb_pl4.*
 - 16: em_ba1.*
 - 17: em_ba2.*
 - 18: em_fun.*
 - 19: em_htgo_hum.*
 - 20: em_htgo_inv.*
 - 21: em_htgo_rod.*
 - 22: em_htg_hum1.*
 - 23: em_htg_hum2.*
 - 24: em_htg_hum3.*
 - 25: em_htg_hum4.*
 - 26: em_htg_hum5.*
 - 27: em_htg_hum6.*
 - 28: em_htg_hum7.*
 - 29: em_htg_hum8.*
 - 30: em_htg_inv1.*
 - 31: em_htg_inv2.*
 - 32: em_htg_other.*
 - 33: em_htg_rod.*
 - 34: em_hum1.*
 - 35: em_hum2.*
 - 36: em_hum3.*
 - 37: em_hum4.*
 - 38: em_hum5.*
 - 39: em_hum6.*
 - 40: em_hum7.*
 - 41: em_in.*
 - 42: em_om.*
 - 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_htg1.*
- 61: gb_htg2.*
- 62: gb_htg3.*
- 63: gb_htg4.*
- 64: gb_htg5.*
- 65: gb_htg6.*
- 66: gb_htg7.*
- 67: gb_htg8.*
- 68: gb_htg9.*
- 69: gb_htg10.*
- 70: gb_htg11.*
- 71: gb_htg12.*
- 72: gb_htg13.*
- 73: gb_htg14.*
- 74: gb_htg15.*
- 75: gb_htg16.*
- 76: gb_htg17.*
- 77: gb_htg18.*
- 78: gb_htg19.*
- 79: gb_htg20.*
- 80: gb_htg21.*
- 81: gb_htg22.*
- 82: gb_htg23.*
- 83: gb_htg24.*
- 84: gb_htg25.*
- 85: gb_pr1.*
- 86: gb_pr2.*
- 87: gb_pr3.*
- 88: gb_pr4.*
- 89: gb_pr5.*
- 90: gb_pr6.*
- 91: gb_pr7.*
- 92: gb_pr8.*
- 93: gb_pr9.*
- 94: gb_ro1.*
- 95: gb_ro2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618	100.0	840	9	AX030177 Sequence
2	606	98.1	11509	1	AE000458 Escherich
3	594	96.1	91414	2	M87049 E. coli gen
4	592.4	95.9	10592	1	AE005613 Escherich
c	557.2	90.2	279589	75	AC073768 Mus muscu
6	486	78.6	212936	66	AC020970 Mus muscu
7	442.6	71.6	96086	3	AF233324 Salmonell
c	415.8	67.3	268294	66	AC020885 Mus muscu
9	339	54.9	290452	76	AC079167 Mus muscu
10	295.8	47.9	2695	2	M30198 E.coli recQ
c	242.8	39.3	62274	66	AC020833 Mus muscu

12	132.8	21.5	298166	67	AC087563	AC087563 Homo sapi
13	120	19.4	4833	78	AC022157	AC022157 Homo sapi
c 14	91.4	14.8	265383	66	AC020874	AC020874 Mus muscu
c 15	88	14.2	10184	1	AE006110	AE006110 Pasteurel
c 16	68.8	11.1	10145	1	AE004719	AE004719 Pseudomon
c 17	68	11.0	11321	1	AE004109	AE004109 Vibrio ch
c 18	66.4	10.7	6246	2	AF235020	AF235020 Brucella
c 19	60.6	9.8	11137	3	U32810	U32810 Haemophilus
c 20	54.8	8.9	672	9	AX030079	AX030079 Sequence
c 21	54.8	8.9	12498	1	AE000140	AE000140 Escherich
c 22	54.2	8.8	4923	3	YEFUABC	Y47200 Yersinia en
c 23	52.4	8.5	128824	2	ECU73857	U73857 Escherichia
c 24	51.8	8.4	12948	1	AE004589	AE004589 Pseudomon
c 25	49.6	8.0	11415	1	AE004889	AE004889 Pseudomon
c 26	48	7.8	2483	3	PACARAB	U04992 Pseudomonas
c 27	48	7.8	7285	3	FAU81259	U13259 Pseudomonas
c 28	47.4	7.7	26245	1	AF087482	AF087482 Pseudomon
c 29	45.4	7.3	294250	2	AP001517	AP001517 Bacillus
c 30	45.2	7.3	10029	1	AE005212	AE005212 Escherich
c 31	45	7.3	10426	1	AE004699	AE004699 Pseudomon
c 32	43	7.0	10976	1	AE004864	AE004864 Pseudomon
c 33	41.8	6.8	792	9	AX063725	AX063725 Sequence
c 34	38.4	6.2	192502	73	AC062017	AC062017 Homo sapi
c 35	37.6	6.1	110000	84	LMFLCHR32_20	Continuation (21 o
c 36	37.4	6.1	159146	75	AC073621	AC073621 Homo sapi
c 37	37	6.0	137137	85	AC004928	AC004928 Homo sapi
c 38	36.8	6.0	6277	5	AF067495	AF067495 Leishmani
c 39	35.6	5.8	3938	93	HUMCD79B	L27587 Human CD79b
c 40	35.6	5.8	11313	1	AE005402	AE005402 Escherich
c 41	35.4	5.7	2271	3	RCHYDFG	Z15087 R.capsulatu
c 42	35	5.7	620	74	AC071237	AX071237 Giardia i
c 43	34.8	5.6	639	9	AX030081	AX030081 Sequence
c 44	34.8	5.6	13793	1	AE000274	AE000274 Escherich
c 45	34.8	5.6	14236	2	D90824	D90824 E.coli geno
ALIGNMENTS						
RESULT 1						
AX030177	AX030177	840 bp	DNA	PAT	16-SEP-2000	
LOCUS	Sequence 3 from Patent	EP1013765.				
DEFINITION	Sequence 3 from Patent	EP1013765.				
ACCESSION	AX030177					
VERSION	AX030177.1	GI:10190394				
KEYWORDS	Escherichia coli.					
SOURCE	Escherichia coli.					
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.					
REFERENCE	1. (bases 1 to 840)					
AUTHORS	Belareva,A.V., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L. and Zakataeva,N.P.					
TITLE	Gene and method for producing l-amino acids					
JOURNAL	Patent: EP 1013765-A 3 28-JUN-2000;					
FEATURES	Location/Qualifiers					
Source	1. 840					
	/organism="Escherichia coli"					
	/db_xref="taxon:562"					
CDS	187..807					
	/note="unnamed protein product"					
	/codon_start=1					
	/transl_table=11					
	/protein_id="CAC09319.1"					
	/db_xref="GI:10190395"					
	/translation="MLMFLTVAMVHIIVAMSPGDFEFVSQTAVSRSRKAMMGVLG					
	ITFCVGMWAGLALGLHLIIIEKMWLHTLIVGGGLYLCWMGIIQMLRALKAEVASP					
	APQVELAKSGFLKPLIKNAWRAIIYFGSVLFLFDVNGVTARGIFALLIIVET					
	LAWEFTVASLFPALPMWRGYYQLAKWIDGAFALFAGFGIHLIISR"					
BASE COUNT	165 a 192 c 248 g 235 t					
BRIGIN						

TITLE
JOURNAL
Direct Submission
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
AUTHORS
3 (bases 1 to 11509)
TITLE
JOURNAL
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
AUTHORS
4 (bases 1 to 11509)
TITLE
JOURNAL
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
On Sep 9, 1997 this sequence version replaced gi:1790254.
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M34. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
source
1. .11509
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
81. .950
/gene="pldA"
/notes="b3821"
81. .950
/gene="pldA"
/EC.number="3.1.1.32"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/notes="0289; 99 pct identical to PA1_ECOLI SW: P00631;
CGSCNo. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/translation="MRTLOGMLLPVEMLPWVAQATVKEVHDPAVRGSIANNLQ
EHNPTFLYPDNYNLIYLTQSDLNKEAISTDWAENARKDEVKQLSLAPLWRGIL
GPNVSITQKSWQOLNSSESPFRETNYEPQLFLGATDYPFAGWTLRDYDMGY
NHDNSGRSDPTSRSNRLYTRLMAENGWLVKVPYVVGNDTDDPDIITKYMGYQLK
IGYHLGDVAVLSAKGYNWNVTGYGGAELGLSYPTIKRVLYTQVYSGYESLIDYNFQ
TRYGVGVMLNDLF"
965. .993
promoter

promoter
1019. .1047
/note="factor Sigma70; predicted +1 start at 4003392"
/note="factor Sigma70; predicted +1 start at 4003446"
1077. .2909
/gene="recQ"
/notes="b3822"
1077. .2909
/gene="recQ"
/EC.number="3.6.1.1"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/notes="0610; 99 pct identical to 607 amino acids
of RECQ_ECOLI SW: PI5043 but has 3 additional N-ter
residues"
/codon_start=1
/transl_table=11
/product="Atp-Dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/translation="MNVAQAEVLNLESGAKOVLOETFGYQQFRPQOEI IDTVLSGRD
CLVVMPTGGKSLCYQIPALLANGLTVVVSPILSLMKDOVDQLOANGVAACLNSTOT
REQLBVMTCGTGQIRLLXIAPERLMDLNFLEHLAHWNPVLLVADEAHCI SONGHDF
RPIYALGQLRQRFPTPLPMATADDTTQDVI RDLGLNDPLIQISSFDRPNIRYM
LMKFKPLDQLMRYVQEQKSGI IVCNSRAKVEDTAARLQSKLSAAAHAGLNNV
RADVQKFRDLDQIVVATVAFGMIKPNPVFVHFDPINIEISYQETGRGDDL
PAMAMLYFPDAMWLRLCLEKPGQGLQIERHKLNAMGAFAPAEQTCRRLVLLNYFG
EGHDEPCGNCIDLPDPKQYDQALSTIGRVNORFGMGYVVEIRGANNQRT
DYGHDKLYVGMGRKSHHWYSVIRLIHLGLVNTQIAQHSALQTEAARPAESS
LQLAVPRIVALPKAKQKSGFNDRKLFKLRKLSIADESNNPPYVVFENDATLIE
MAQMPITASEMLSVNGVGMRKLERFPFMAIIRAHVDDGDEE"
2104. .2140
repeat_region
/notes="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
3162. .3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224. .3592
/gene="yigJ"
/notes="b3823"
3224. .3592
/gene="yigJ"
/function="orf; Unknown"
/note="0322; sequence change shortens and
changes N-terminus relative to earlier version (YIGJ_ECOLI
SW: P27846)"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76826.1"
/db_xref="GI:2367302"
/translation="MGYQMLRGALKKEAVSAPQOVELAKSGRSFLKGLLNLNPKA
IIFYGVSFLFYGDNVGTTRWGIFALIIVETLAWFTVVASLFPALPQMRRGYQRLAKW
IDGFAGALFAGFGIHLIISR"
3612. .3645
repeat_region
/notes="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
complement(3654. .4070)
/gene="yigK"
/notes="b3824"
complement(3654. .4070)
/gene="yigK"
/function="orf; Unknown"
/note="f138; 100 pct identical to YIGK_ECOLI SW: P27847"
/transl_table=11
/codon_start=1
/product="orf, hypothetical protein"
/protein_id="AAC76827.1"
/db_xref="GI:1790258"
/translation="MIAFVFLKWAAGAAVLIWLGIQOWRAAGAI DLKSLASTQSRRLP
QRAVFNLINPKSIVFLAALFPQFIMPOQPQLMQYIVLGVTIVVDIIVMIGIYTLAQ
RIALWIKGPKQKALNKIFGSLFMLVGLALLASARHA"
4200. .4230
promoter
/note="factor Sigma70; predicted +1 start at 4006629"
complement(4227. .4254)
promoter

/note="factor Sigma70; predicted +1 start at 4006612"
4384..5406
/gene="pldB"
/note="b3825"
4384..5406
/gene="pldB"
/EC_number="3.1.1.5"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/note="o340; 99 pct identical amino acid sequence and
equal length to PLDB_ECOLI SW: P07000"
/codon_start=1
/transL_table=11
/product="lysophospholipase L(2)"
/protein_id="AAC76828.1"
/db_xref="GI:2367303"
/translation="MFQOKDWETRENAFAAFTMGPLTDFWRQDEAETGVDIDPVR
FYFRQAHDRVVICPGRIESYKVAELAYDLFLGFDVLIIDHRGGRSGRLIADP
HGHVNRFDYDDDLAAFWQEQVQPGPKRYILAHSMGSAISTLFQRHPGVCDIA
LTAPMFIYIRMPFMAWQILNMAEAPRFDGYAIGTGRWALPFAINVLHRSQRY
RRLRFYADDPITRVGGPTYHWRRESILAGEQVLGAGDDATPTLLQAEERVDNR
MDRFCELRTAAGHPVEGGRPLVIKAYHEILFEKDMASVALHAIYDFNHRNSPSG
NRSTEV"
5381..5408
/note="factor Sigma70; predicted +1 start at 4007807"
5426..5455
/note="factor Sigma70; predicted +1 start at 4007854"
5438..5466
/note="factor Sigma70; predicted +1 start at 4007865"
5494..5523
/note="factor Sigma70; predicted +1 start at 4007922"
5526..6041
/gene="yigL"
Query Match 98.1%; Score 606; DB 1; Length 11509;
Best Local Similarity 99.8%; Pred. No. 3.9e-167;
Matches 617; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 atgttgatgttattctacocgtgcgatgtgcacattgtggcgcttatgagcccggt 60
Db 2973 ATGTTGATGTTATTCTCACCGTCGCCATGTGTGCACATGTGGCGCTTATGAGCCCGGT 3032
Qy 61 ccgattctttttgtctccacgcgtgtcagtcgtccgttaacgaagcgcgatgtg 120
Db 3033 CCCGATTTCTTTTGTCTTCACACCCCTGTCTAGCTGCTCCGTAAAGACGATGATG 3092
Qy 121 ggcgtgctggcattacacgcgcgtaagtgttggtggcggtgattgcgctgtggcctg 180
Db 3093 GCGGTGCTGGCAATTACCTGCGCGCTAATGTTTGGGCTGGGATTCGCTGCTGGCCTG 3152
Qy 181 catttgattatcgaaaaatggcctggctgcatacgtgattatgttgccggtgacctg 240
Db 3153 CATTGTGATATCGAAAAATGGCTGGCTGCATACGCTGATATGTTGGCGGTGG-CTG 3211
Qy 241 tatctctgtagtggttaccagatgctacgtgtgctgactgaaagagcggtttct 300
Db 3212 TATCTCTGCTGGAGGGTTACAGATGCTACGTGTGCACACAAAAGAGCGGTTCCT 3271
Qy 301 gcacctgcccacaggtgcagctggcgaaaagtgggcgagtttccctgaaaggttactg 360
Db 3272 GCACCTGCGCCACAGTGCAGTGGCGAAAAGTGGCGCAGTTTCTGAAAGGTTTACTG 3331
Qy 361 accaatctcgtaatccgaagcgaattatctacttgcctcggtgtctcattgtttgct 420
Db 3332 ACCAATCTCGTAAATCCGAAGCGAATTATCTACTTTGGCTCGGTGTTCTCATGTTGTCT 3391
Qy 421 ggtgataacgttggcactaccgcgcgctggggcatttttgcgctgatacattgtcgaagc 480
Db 3392 GGTGATACGTTGGCAGTACCGCGCGCTGGGCAATTTTTCGCTGATCATTTGCGAAACG 3451
Qy 481 ctggcggtggtttaccgctgcgttgcagcgcgtgttttgcctgcgcgaatgcgcggtggttat 540
Db 3452 CTGGCGGTGTTTACCGCTGTTGCCAGCCTGTTTGGCCCTGCCCAATGCCCGCTGGTTAT

Qy 541 caacgtctgcgaagtgattgatgttttgcggcggttatttgcggtatttgcggtatttgcatt 600
Db 3512 CAACGCTGCGGAAGTGGATGATGTTTTCGCGGGCGTATTATTCGCGGATTTGGCATT 3571
Qy 601 catttgattatttcgcgg 618
Db 3572 CATTGATTATTTCGCGG 3589
RESULT 3
LOCUS ECOUW85 91414 bp DNA BCT 29-MAY-1995
DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
ACCESSION M87049
VERSION M87049.1 GI:836656
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 91414)
Daniels,D.L., Plunkett,G. III., Burland,V. and Blattner,F.R.
Analysis of the Escherichia coli genome: DNA sequence of the region
from 84.5 to 86.5 minutes
Science 257 (5071), 771-778 (1992)
92358234
2 (bases 1 to 91414)
Plunkett,G. III., Burland,V., Daniels,D.L. and Blattner,F.R.
Analysis of the Escherichia coli genome. III. DNA sequence of the
region from 87.2 to 89.2 minutes
Nucleic Acids Res. 21 (15), 3391-3398 (1993)
93347969
3 (bases 1 to 91414)
Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and
Daniels,D.L.
Analysis of the Escherichia coli genome. IV. DNA sequence of the
region from 89.2 to 92.8 minutes
Nucleic Acids Res. 21 (23), 5408-5417 (1993)
94089392
4 (bases 1 to 91414)
Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and
Rouviere,P.E.
A new family of peptidyl-prolyl isomerases
Trends Biochem. Sci. 20 (1), 12-14 (1995)
95184296
5 (bases 1 to 91414)
Daniels,D.L.
Direct Submission
Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax:
608-263-7459
On May 29, 1995 this sequence version replaced gi:148169.
This sequence was determined as part of the E. coli Genome Project
(Frederick R. Blattner, director) at the University of
Wisconsin-Madison. The entire sequence was independently determined
from E. coli MG1665. Overlaps and conflicts with other sequence
determinations are annotated. Reference [1] describes the original
sequence determination of a 91408 bp sequence. References [2], [3],
and [4] describe subsequent corrections and/or updates to that
sequence.
FEATURES
Location/Qualifiers
1. 91414
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/clone_lib="lambda library of Daniels and Blattner"
/map="bp 1 at 3974 kb; 84.7 minutes"
/note="This sequence comprises the following lambda
clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177),
DD856(EC19RM8.1), DD858(EC17RM9),

DD861 (EC17-221), DD864 (EC27-149), DD865 (EC21-76),
DD867 (EC27-860). M13mp19 or Janus vectors were used for
subcloning."

1..200
/note="corresponds to K02846 (1..200)"
1..5863
/note="clone EC15-116"
complement(1..4899)
/note="corresponds to M37337 (1..4900)"
1..339
/note="corresponds to V00349 (164..500)"
10..86
/gene="aspT"
/note="CGSC No. 989"
10..86
/gene="aspT"
95..170
/gene="trpT"
/note="CGSC No. 66"
95..170
/gene="trpT"
177..205
/note="terminator-like sequence"
253..396
/note="predicted bend of 81 degrees"
complement(266..862)
/gene="fl98"
complement(266..862)
/gene="fl98"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67566.1"
/db_xref="GI:148170"
/translation="WAHTSRHNEFSIGASALWECMLNOWLGRLYONQDAHTGLQFEA
RIAQKOSLVQLHERQLDLITTEAPKMFDEFSSQLIGYFTLALTSAPSKLKGDNLNL
RLEWPDFQOHEAGLEAGDEVPIITTSAAELAQOQIAMLNGCTWLPVSWARKKGLHT
VVDSTLSRPLAYIWLQNSDKNALIRDLKINVLDEV"
586..724
/note="predicted bend of 77 degrees"
complement(703..1104)
/gene="fl33"
complement(703..1104)
/gene="fl33"
/note="CGSC No. 18010; possible pssr gene"
/codon_start=1
/transl_table=11
/protein_id="AAA67567.1"
/db_xref="GI:148171"
/translation="MDELLKTFLEVSRHFGRAESILYLTOSAVSFRIQLENQLG
VNLFRHRNIRLTAAGEKLLPYAETLMTSQAAKRWIRPHDITSLFSLVPAPRCGNV
CLISGWDACIKRMPIQAYSKRELPLNGSLW"
complement(1136..1165)
1151..1179
/note="promoter matrix score of 63"
/note="promoter matrix score of 57"
1223..1636
/gene="ol37"
1223..1636
/gene="ol37"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67568.1"
/db_xref="GI:148172"
/translation="MAESFTTNRFDNKHYPGRGFSRHGDDTIKEAOILLERGHYAFNE
LDLGRKPEVTEEEKLFVAVCGEREPVTEAERVMSKYMTRIKRPRKREHTLSAVNRRLK
VLKTPILTINKGEMPFYVVSFCRWMSKRSIAR"
1562..1586
/gene="ol37"
/note="terminator-like sequence"
complement(1585..3135)
/gene="f516"

CDS
complement(1585..3135)
/gene="f516"
/note="similar to Mg chelatase subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67569.1"
/db_xref="GI:148173"
/translation="MHIMVTOGLMSLSIVHTRAAAGVGNAPITVEVHISKGLPLTM
VGLPETTVKEARDVRVSALINSGYEYPAKTIINLAPDLPEKGRVDLPITAIALLA
SEOLTANKLDSEYELVGEALGALGVPCALSSATEAIKSKKIIVAKNDDEVGLLA
GEGCLLIADHLQAVCAFLEGGKHALEKPTDAVSRAHQDLSDVIGQGGKRGLEITAA
GHNLLIIGPPTGKTMLASINGLLPDLNSNEALESAAIILSVNAESVQKQRPFP
RSRPHSASLTAMVGGAIPGPGEISLAHNGVLFDELPEFERRTLDALRPTIESGQTH
LSRTRAKITYPARFQLVAAAMPSPTHVQGNHNRCTPEOTLRYLNRLSGPFLODFDL
LEIPLPPGILSKTVVPGESSATVKQVMAAREQKRONKLNALDPSPEIQPCKLE
SEDAMWLEGTLIHLGLSIRAWQRLKLVARTIADIQSDITTRQHLQEAIVSTRAIDRL
IHLQKLLT"
3287..3433
/note="predicted bend of 79 degrees"
3297..3326
/note="promoter matrix score of 74"
3336..3676
/note="corresponds to M11651 (1..341)"
3336..3737
/note="corresponds to V00289 (1..402)"
3336..12791
/note="corresponds to M10313 (1..9456)"
3379..5721
/note="V00290; ECILVX(1..2343)"
3458..3556
/gene="ilvL"
3458..3556
/gene="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGEDA operon leader peptide"
/protein_id="AAA67570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVISVVVVIIPCGAALGRKA"
3696..4679
/gene="ilvG"
3696..4679
/gene="ilvG"
/note="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetohydroxy acid synthase II, large subunit"
/protein_id="AAA67571.1"
/db_xref="GI:148175"
/translation="MNGAOWVHALRAQGVTVFGVGGAIMPVYDALYDGGVHEHLC
RHEQGAAMAAGVARATGKVCATSGCATNLITGLADALLDSIPVVAITQVNSAP
FIGTDAFQVDYDLGLSLACHSFLVQSLEELPRIMAEAFDVACSGRPGVLDIPKD
IQLASGLDEFWFTTVENAVFPHAEVQAKMLAKPMLYVGGVGMQAQVAPLRE
FLAATKMPATPCTILKGLGAEADYPYILGMLGMHGTKAANFAVQECDLLIAVGAREDDR
VTGLNTFAPHASVIHMDIDPAEMNKLRQAHVALQGLNALLPALQOPLNQ"
4675..5340
/gene="o221"
4675..5340
/gene="o221"
/note="the mutation ilv02096 joins this ORF to ilvG,
restoring function missing in wild-type"
/pseudo
/codon_start=1
replace(4677..4678,"tatg")
variation
Query Match 96.1%; Score 594; DB 2; Length 91414;
Best Local Similarity 99.7%; Pred. No. 1.5e-163;
Matches 616; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 atgttgatgtattttccacgcgtgcacattgtgcacattgtgcgcttatgaccccggt 60
|||||

Db	60863	ATGTTGATGTTATTTCTCACCGTCGCCATCGTGTGCACATTTGGCGCTTATGAGCCCGGT	60922
Qy	61	ccgattctcttttttctctcagacgcgtgctcagtgctcccgtaagaagcgatgatg	120
Db	60923	CCCGATTTCTTTTGTCTCTCAGACCGCTGTCAAGTCGTTCCCTAAGAAGCGATGATG	60982
Qy	121	ggcgtgctggcattacctgcggcgtaatagggtttggcgctgggattgcgctcttggcctg	180
Db	60983	GGCGTGTGGGCATTTACCTGGCGGCTAATGGTTTGGCGTGGGATGGCGTCTTGGCGCT	61042
Qy	181	catttgattatcgaaaaagcctgctgcatacgcctgattatggcggtggcgctg	240
Db	61043	CATTTGATATCGAAAAAAGGCCCTGGCTGCTATACGCTGATTATGGTGGCGGTGG-CTG	61101
Qy	241	tatctctgctggatgggtttaccagatgctacgtggtgcactgaaataagagcggtttct	300
Db	61102	TATCTCTGCTGGATGGGTTACACAGATGCTACGTGTGTGCACATGAAAAAAGAGCGGTTTCT	61161
Qy	301	gcacctgcgcacaggctcagctcggcgaaataagtgggcgagcttctcgaaggtttactg	360
Db	61162	GCACCTGCGCCACAGGTGAGCTGCGGAAAGTGGGCGACGTTTCCCTGAAA-GTTTACTG	61220
Qy	361	accaatctcgttaatccgaaagcagatctatctactttgctcggtggttctcattgtttgc	420
Db	61221	ACCAATCTCGTATATCCGAAAGCGATATCTACTTTGCTCGGTGTTCTCATGTTTGTCTC	61280
Qy	421	ggtgataacgttggcactaccgcgcgtcggtgggcatatttgcgtgcatcattgcgaaacg	480
Db	61281	GGTGATAACGTTGGCACATACCgcgcgcgtGGGCGCATTTTGGCGCTGATCATTTGCGAAACG	61340
Qy	481	ctggcgtggtttacgcgtgttgcagcctggtttccctgcgcgaataatgcgcgtggttat	540
Db	61341	CTGGCGTGGTTTACCGTGGTTGCGCAGCTGTTTGGCCCTGCGCAAAATGCGCGTGGTTAT	61400
Qy	541	caacgctgcgcgaagtgattgatggttttgcggggcggttatttgcgcgatttggcatt	600
Db	61401	CAACGCTGCGCGAAGTGGATGATGTTTGGCGGGCGGTTATTTGCCGGATTGGCATT	61460
Qy	601	catttgattatttcgcgg	618
Db	61461	CATTGATGATTATTCGCGG	61478
RESULT 4			
AE005613			
LOCUS	AE005613	10592 bp	DNA BCT 25-JAN-2001
DEFINITION	Escherichia coli O157:H7 genome, contig 3 of 3, section 232 of 290.		
ACCESSION	AE005613 AE005174		
VERSION	AE005613.1 GI:12518685		
KEYWORDS	.		
SOURCE	Escherichia coli O157:H7.		
ORGANISM	Escherichia coli O157:H7. Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.		
REFERENCE	1 (bases 1 to 10592) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		
TITLE	Genome sequence of enterohemorrhagic Escherichia coli O157:H7		
JOURNAL	Nature 409 (6819), 529-533 (2001)		
REFERENCE	2 (bases 1 to 10592) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.		
TITLE	Direct Submission		
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA		

Location/Qualifiers	
1..10592	/organism="Escherichia coli O157:H7"
	/strain="EDL933"
	/serotype="O157:H7"
	/db_xref="taxon:83334"
	/note="enterohemorrhagic"
<1..71	
	/note="O-island #155; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
	complement(72..836)
	/gene="z5332"
	complement(72..836)
	/gene="z5332"
	/function="orf; Unknown function"
	/note="Residues 1 to 131 of 254 are 95.41 pct identical to residues 1 to 131 of 161 from Escherichia coli K-12 Strain MG1655; B3815"
	/codon_start=1
	/transl_table=11
	/protein_id="AAG59008.1"
	/db_xref="GI:12518686"
	/translation="MAHRLIIGKGMITLNLKRIFLALTLLPLFAVAADDCALSDPILT VQATYVPQTERVKMYMQKANGAWGLTALLADMSQGVQVMAMNGGIYDESTAPLG LYTENGOQVALNLAGSEGNFIRPQGVFFVAGDKVIVRLDAFKTSKEIQFAVQSGP MLMENSVINPRIHPNVASRKIRNGVGINKHGNVFLLSQOATNFYDPACYAKAKLNVE OLLYLDGTSISHMYMKGAIPWQYFPFVTMISVERKG"
	1206..2156
	/gene="corA"
	/note="z5333"
	1206..2156
	/gene="corA"
	/function="transport; Transport of small molecules; Cations"
	/note="Residues 1 to 316 of 316 are 100.00 pct identical to residues 1 to 316 of 316 from Escherichia coli K-12 strain MG1655; B3816"
	/codon_start=1
	/transl_table=11
	/product="Wg2+ transport, system I"
	/protein_id="AAG59009.1"
	/db_xref="GI:12518687"
	/translation="MLSAFQLENRLRLEVEESQPLVNAVWIDLVEPDDDERLRVOS ELGQSLATRPELEDIEASARFFEDDDGLHIHFFFEADAEDHAGNSVAFITRGRLLF TRERELPAFLRYMRARSQSMVDGNAYELILDIFETKIEQLADEIENIYDLEQLSR VIMSGHGDVEYDEALSTLAELEDIGKVRCLMDTQALNLFVRKARLPGQLQOARE ILRDIESLLPHNESLFQKVNFLMQAAMGFINEONRIKIFSVVSVFLPPTLVASSY GMNEFPELPKLKWSFGYGAIFMILAGLAPLYLFRKNWL"
	2157..4848
	/note="O-island #156; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
	complement(2198..2689)
	/gene="z5334"
	complement(2198..2689)
	/gene="z5334"
	/function="orf; Unknown function"
	/note="No significant matches"
	/codon_start=1
	/transl_table=11
	/protein_id="AAG59010.1"
	/db_xref="GI:12518688"
	/translation="MKSKLQKLIILCLFLCCIYNLWLRPVQILYTYSDAGNSVFLVV DHLPTWDSDKINWYLKHQNEIKNHPLEGSNHWYVIDINGFTDYKKIIEGPEDL YCFPTIKSNDNCITKNYLMWINEYPYRNTHIGINDTEYQLTQENKIEQVFNPHNFK DNF"
	complement(2686..3534)
	/gene="z5335"
	complement(2686..3534)
	/gene="z5335"
	/function="orf; Unknown function"
	/note="No significant matches"
	/codon_start=1
	/transl_table=11

```
gene
  /protein_id="AAG59011.1"
  /db_xref="GI:12518689"
  /translation="MNTLLPSINIPCTLFETISLFDNYSADDMQYDMVEQDFLSLG
  LSDISAKVYRLIKHFPGPGINVAFSASSGTRKISQRECTDILFAEMKELAKMFS
  FPGQYTLIEDLIEHRYGNFNHSHQQLNLSFHEKINKYGNSPIRIKKEICNGIN
  STPGYQPLQISIKTLKLISSLNKFNDFSNGLGISVHDI SAOKI SLISFOKYA
  IGWSATIFHAQDHGLDVTIDKNKTKSYKRFRIWFLQRHKDEAFKPFPTNFNTIE
  RIENYL"
  3748..3867
  /gene="Z5336"
  CDS
  3748..3867
  /gene="Z5336"
  /function="orf; Unknown function"
  /note="Residues 1 to 39 of 39 are 100.00 pct identical to
  residues 278 to 316 of 316 from GenPept 118 :
  gi16960250|gb|AAAF33440.1| (AF233324) S. typhimurium
  magnesium transport protein (COR) (SP:P31138) [Salmonella
  typhimurium LT2]"
  /codon_start=1
  /transl_table=11
  /product="partial repeat of corA"
  /protein_id="AAG59012.1"
  /db_xref="GI:12518690"
  /translations="MNFEMPELKWSEGYGALIFMILAGLAPLYLFRKKNWL"
  complement(3892..4278)
  /gene="Z5337"
  CDS
  complement(3892..4278)
  /gene="Z5337"
  /function="orf; Unknown function"
  /note="No significant matches"
  /codon_start=1
  /transl_table=11
  /protein_id="AAG59013.1"
  /db_xref="GI:12518691"
  /translation="MFKSSIRNIFPSKLERLSRFQFIYITLTSWSILITLNDINLAL
  SFILVNNTVINIGAKIRIDTGCRAIYFVLISGLTLFSIIAIEGVRGESLSVQ
  LNVAILFTTIFMLSAIPSAKMSLSAH"
  4483..4794
  /gene="Z5339"
  CDS
  4483..4794
  /gene="Z5339"
  /function="orf; Unknown function"
  /note="No significant matches"
  /codon_start=1
  /transl_table=11
  /protein_id="AAG59014.1"
  /db_xref="GI:12518692"
  /translation="MDSIEMVTFGSCSWKNPYKKSDFKVKNLLWNAITILVLFIL
  SLITTLMLRSIGLSFGFCQILGRYTMISLGSWVGLGFPFVNI VLFLLIFPKK
  R"
  complement(4841..5728)
  /gene="Z5340"
  CDS
  complement(4841..5728)
  /gene="Z5340"
  /function="orf; Unknown function"
  /note="Residues 4 to 295 of 295 are 92.12 pct identical to
  residues 8 to 299 of 300 from Escherichia coli K-12 Strain
  MG1655; B3820"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAG59015.1"
  /db_xref="GI:12518693"
  /translation="MDAKTRQCVLLAALYFIWGIAPAFYKLIYVPADEILTHRVI
  WSPFFVMVLSICRWSYIKTIQTQPKIFMLAVSAVLGGNLLFIWVNNHHMLEA
  SLGYFIPNLVIMISICRWSYIKTIQTQPKIFMLAVSAVLGGNLLFIWVNNHHLEA
  FARYGLVKKIAYEAOITGMLETMMLPVAATYVLPALADSPSHMGONPMSLNLLIIA
  AGIVTVPLCLCFCAATRLRLSLTGFQVIGTFLFLLAVTYGKPGADRKMTFTFI
  WVALAFVMDAILTQRKS"
  complement(5780..6265)
  /gene="yigI"
  /note="Z5341"
  gene
  complement(5780..6265)
  /gene="yigI"
  /function="orf; Unknown function"
  /note="Residues 1 to 161 of 161 are 100.00 pct identical
  to residues 1 to 161 of 161 from Escherichia coli K-12
  Strain MG1655; B3820"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAG59016.1"
  /db_xref="GI:12518694"
  /translation="MKCKDDMSAVLTAEQALKLVGEMFYHMPFNRAIGMELEYEKE
  FAQLAFKNQPMVMVWNAQSLILHGCVIAGLVAGVSTLTHHETISDELQRRL
  SRMGTTDLRVYDLRPGRXERFATSTSLLRAGNKVAVARVELHNEEQYLIASATATYVW
  G"
  6412..7281
  /gene="pIdA"
  /note="Z5342"
  6412..7281
  /gene="pIdA"
  /function="enzyme; Central intermediary metabolism:
  Phosphorus compounds"
  /note="Residues 1 to 289 of 289 are 100.00 pct identical
  to residues 1 to 289 of 289 from Escherichia coli K-12
  Strain MG1655; B3821"
  /codon_start=1
  Query Match 95.9%; Score 592.4; DB 1; Length 10592;
  Best Local Similarity 97.4%; Pred. No. 3.9e-163;
  Matches 602; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
  QY 1 atgtgtatgtatttctcacgcgtgcgcattgtgcacattgtggcgcttatgagcccccgtt 60
  Db 9307 ATGTGTATGTATTTCACCGTCGCCATGGTGCACATTGTGGCGCTTATGAGTCCCGGT 9366
  QY 61 cccgattctttttgtctctcagaccgctgtcagtcgctccgtaaaagacgatgatg 120
  Db 9367 CCCGATTCTTTTGTCTCTCAGACGCGTGTGAGTCGTCGCCGTAACAGCGATGATG 9426
  QY 121 ggcgtgctggcgattacctgcgcgtaagtgtttggcgctgggatgcgctgcttggcctg 180
  Db 9427 GGCCTGTGGGTATTACCTCGCGCGTTATGGTTTGGCGAGGATTGCGTTGGCGCTG 9486
  QY 181 catttgattatcgaaaaatggcctggctgcatacgtcattgattgctggcggtgctgctg 240
  Db 9487 CATTTGATTATCGAAAAAATGGCTGGCTGCATACGCTGATTTATGTTGCGGTGGCCTG 9546
  QY 241 tatctctgctggatgggtaccagatgctacgtggtgcactgaaaaaagagcggtttct 300
  Db 9547 TATCTCTGCTGATGGGTATTACAGATGCTGCGCGGCGCGCTGAAAAAAGAGCGGTTCT 9606
  QY 301 gcacctgcgcacaggctcagctggcgaaaaagtggcgagtttctctgaaagtttactg 360
  Db 9607 GCACCTGGCCACAGCTCGAGCTGGCGGCGGCGAGTTTCTCTGAAAGGTTTACTG 9666
  QY 361 accaatctcgtaataccgaagcattatctactcttggctcggtgttctcattgtttgtc 420
  Db 9667 ACCAATCTCGTAATCGGAAAGCGATTATCTACTTGGCTCGGTGTTCTCAITGTTGTC 9726
  QY 421 ggtataacgtttgcacacgcgcgctggcgatttttgcgctgatactatcgaacg 480
  Db 9727 GGTATACGTTTGCACCTACCGCGCGCTGGGCGATTTTTCGGTTGATTATTTCGAAACG 9786
  QY 481 ctggcgtggtttaccctcgctgtgcagcctgtttgcctgcgcgaatgcgcgctggttat 540
  Db 9787 CTGGCGTGGTTTACTGTGTGGCGCAGCGCTGTTTGGCGCTGCGCAATGCGCGGTGTTAT 9846
  QY 541 caacgtctgcgaagtgatgattgttttgcggcgacatttatttcggatttgcatt 600
  Db 9847 CAACGTCGGCGAAGTGATGATGATTTTTCGGCGGCGGTATTTCGCCGATTTGCGCAT 9906
  QY 601 catttgattatttcgcgg 618
  ||||||||||||||||
```

```

Db 9907 CATTTCATTATTCGGG 9924
RESULT 5
AC073768/c
LOCUS AC073768.1 GI:8810385
DEFINITION Mus musculus clone Rp23-359N18, WORKING DRAFT SEQUENCE, 60
ACCESSION AC073768
VERSION AC073768.1
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1875058
Center clone name: RPCI-23_359N18
-----
Summary Statistics
Consensus quality: 230709 bases at least Q40
Consensus quality: 253950 bases at least Q30
Consensus quality: 258916 bases at least Q20
Estimated insert size: 216000; agarose-fp estimation
Quality coverage: 7.54 in Q20 bases; agarose-fp estimation
Quality coverage: 9.53 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1009: contig of 1009 bp in length
* 1010: gap of unknown length
* 1109: contig of 1023 bp in length
* 2132: gap of unknown length
* 2133: contig of 1025 bp in length
* 2233: contig of 1025 bp in length
* 3257: gap of unknown length
* 3358: 4550: contig of 1193 bp in length
* 4651: 4650: gap of unknown length
* 6414: contig of 1764 bp in length
* 6415: 6514: gap of unknown length
* 6515: 7557: contig of 1043 bp in length
* 7558: 8885: contig of 1228 bp in length
* 7658: 8985: gap of unknown length
* 8886: 10366: contig of 1381 bp in length
* 8986: 10466: gap of unknown length
* 10367: 11691: contig of 1225 bp in length
* 10467: 11791: gap of unknown length
* 11692: 13116: contig of 1325 bp in length
* 11792: 13216: gap of unknown length
* 13117: 14234: contig of 1018 bp in length
* 13217: 14334: gap of unknown length
* 14235: 15472: contig of 1138 bp in length
* 14335: 15572: gap of unknown length
* 15473:

```

```

* 15573: contig of 1117 bp in length
* 16690: gap of unknown length
* 16789: contig of 1049 bp in length
* 17838: gap of unknown length
* 17939: contig of 1250 bp in length
* 19188: gap of unknown length
* 19288: gap of unknown length
* 20826: contig of 1538 bp in length
* 20926: gap of unknown length
* 21984: contig of 1058 bp in length
* 22084: gap of unknown length
* 23158: contig of 1074 bp in length
* 23159: gap of unknown length
* 23258: contig of 1339 bp in length
* 23259: gap of unknown length
* 24598: gap of unknown length
* 24598: contig of 1087 bp in length
* 25784: gap of unknown length
* 25785: contig of 1045 bp in length
* 26929: gap of unknown length
* 26930: gap of unknown length
* 27030: contig of 1506 bp in length
* 28535: gap of unknown length
* 28536: contig of 1393 bp in length
* 30028: gap of unknown length
* 30228: gap of unknown length
* 31331: contig of 1203 bp in length
* 31331: gap of unknown length
* 31431: contig of 1080 bp in length
* 31432: gap of unknown length
* 32512: gap of unknown length
* 32612: contig of 1278 bp in length
* 33889: gap of unknown length
* 33890: contig of 1203 bp in length
* 35192: gap of unknown length
* 35192: contig of 1203 bp in length
* 35292: gap of unknown length
* 36574: contig of 1282 bp in length
* 36574: gap of unknown length
* 36675: contig of 1270 bp in length
* 37945: gap of unknown length
* 38044: contig of 1085 bp in length
* 39129: gap of unknown length
* 39229: gap of unknown length
* 40571: contig of 1342 bp in length
* 40572: gap of unknown length
* 42172: contig of 1501 bp in length
* 42173: gap of unknown length
* 42273: contig of 1895 bp in length
* 44168: gap of unknown length
* 44267: gap of unknown length
* 45266: contig of 1459 bp in length
* 45268: gap of unknown length
* 45727: contig of 1663 bp in length
* 45827: gap of unknown length
* 47490: gap of unknown length
* 47590: contig of 1289 bp in length
* 48878: gap of unknown length
* 48879: contig of 1788 bp in length
* 48979: gap of unknown length
* 50766: contig of 3371 bp in length
* 50767: gap of unknown length
* 50866: contig of 2183 bp in length
* 54237: gap of unknown length
* 54238: contig of 1513 bp in length
* 56201: gap of unknown length
* 56221: contig of 1294 bp in length
* 58134: gap of unknown length
* 58233: contig of 3840 bp in length
* 59527: gap of unknown length
* 59627: contig of 2160 bp in length
* 63468: gap of unknown length
* 63468: contig of 2160 bp in length
* 63568: gap of unknown length
* 65728: contig of 2960 bp in length
* 65728: gap of unknown length
* 68787: gap of unknown length
* 68788: contig of 3429 bp in length
* 68888: gap of unknown length
* 72316: contig of 4159 bp in length
* 72317: gap of unknown length
* 76575: gap of unknown length
* 76576: contig of 5597 bp in length
* 76676: gap of unknown length
* 82273: contig of 3827 bp in length
* 82373: gap of unknown length
* 86299: gap of unknown length
* 86300: contig of 5547 bp in length

```

```
* 91847 91946: gap of unknown length
* 91947 96394: contig of 4448 bp in length
* 96395 96494: gap of unknown length
* 96495 102483: contig of 5989 bp in length
* 102484 102583: gap of unknown length
* 102584 110476: contig of 7893 bp in length
* 110477 110576: gap of unknown length
* 110577 119241: contig of 8665 bp in length
* 119242 119341: gap of unknown length
* 119342 130059: contig of 10718 bp in length
* 130060 130159: gap of unknown length
* 130160 142204: contig of 12045 bp in length
* 142205 142304: gap of unknown length
* 142305 157417: contig of 15113 bp in length
* 157418 157517: gap of unknown length
* 157518 185392: contig of 27875 bp in length
* 185393 185492: gap of unknown length
* 185493 212309: contig of 26817 bp in length
* 212310 212409: gap of unknown length
* 212410 249959: contig of 37550 bp in length
* 249960 250059: gap of unknown length
* 250060 279589: contig of 29530 bp in length.
FEATURES
    Location/Qualifiers
        1..279589
            /organism="Mus musculus"
            /db_xref="taxon:10090"
            /clone_lib="RP23-359N18"
            /clone_lib="RPC1 mouse BAC library 23"
BASE COUNT 69254 a 68603 c 67231 g 68544 t 5957 others
ORIGIN

Query Match      90.2%; Score 557.2; DB 75; Length 279589;
Best Local Similarity 98.7%; Pred. No. 1..1e-152;
Matches 614; Conservative 0; Mismatches 3; Indels 5; Gaps 5;

QY 1 atgttgatgtatttctcaacgtcgccatggtgcacattgtggccttatagcccggt 60
    |||||
Db 30830 ATGTGATGTTATTTCTACCGTCGCCATGTGACATGTGGCGTATGAGCCCGGT 30771
    |||||

QY 61 cccgatttcttttctctcagaccgctgcagtcagtcgcttccgttaagaagcagatg 120
    |||||
Db 30770 CCGATTTCTTTTGTCTCTCAGACCCTGTCAGTCGTCCGCTAAAGAAGCGATG 30711
    |||||

QY 121 ggcgtgctggcattacacctgcggcgtaagtgttggcctgggattgcctctggcctg 180
    |||||
Db 30710 GCGTGTCTGGCATTAACCTGCGCGCGTAAATGTTGGCTGGGATGCGCTGTGGCCTG 30651
    |||||

QY 181 catttgattatcgaaaaaattggcctggcgtgcatacgcgtgattatggtggcggtggcctg 240
    |||||
Db 30650 CATTTGATTATCGAAAAAATGGCCTGGCTGCATACGCTGATTATGTTGGCGGTGGCCTG 30591
    |||||

QY 241 tatctctgctgagtggttaccagatgctacgtggtgcactgaaagagag-cgggttc 299
    |||||
Db 30590 TATCTCTGCTGATGGGTTTACCAGATGCTACGTGGTGGCTGAAAGAGCGCGGTTTC 30531
    |||||

QY 300 tgcacctgcgcacaggtgcagctggcgaaagtggcgagtttccctgaaagtgtttact 359
    |||||
Db 30530 TGCACCTGCGCCACAGTGCAGCTGGCGAAGAGTGGCGCAGTTTCTCTGAAAGGTTTACT 30471
    |||||

QY 360 gaccatctcgctaatccgaaagcattatctacttggctcgatgtctctcattgtttgt 419
    |||||
Db 30470 GACCATCTGCTGATTCGGAAGAGCATATATCTTTCGCTGGGTGTTCTCATTTGTTGT 30411
    |||||

QY 420 cgggtgataacgttggcactaccgcgcgtggggcatttttgcctgatca-ttgcgaaa 478
    |||||
Db 30410 CGGTGATAACGTTGGCAC"TAACCGCGCGTGGGGCATTTTTCGCTGTATCACTTTCGAAA 30351
    |||||

QY 479 cgtgcgctgggtttaccgctcgtt-gccagcctgt-ttgcctgcgcgaaatgcgccgtgg 536
    |||||
Db 30350 CGCTGCGGTGGTTTACCGTCTGTCGCCAGCCGTCTTGCCTTCGCCAAATGCGCCGTGG 30291
    |||||

QY 537 ttatcaacgctctggcgaaagtggattgatgtgttttgcggggcggttatttgcgggattgg 596
    |||||
```

```
Db 30290 TTATC-ACGTCTGGCAAGTGGATTGATGTTTTCGCGGCGGTATTTCGCGGATTGG 30232
    |||||
QY 597 cattcatttgattatttcgcgg 618
    |||||
Db 30231 CATTCATTGATTATTTCGCGG 30210
    |||||

RESULT 6
AC020970 212936 bp DNA HTG 10-FEB-2000
LOCUS Mus musculus clone RP23-252M21, WORKING DRAFT SEQUENCE, 144
DEFINITION unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
* 11047 12223: contig of 1177 bp in length
* gap of unknown length
* 12224 13492: contig of 1269 bp in length
* gap of unknown length
* 13493 14669: contig of 1177 bp in length
* gap of unknown length
* 14670 15886: contig of 1217 bp in length
* gap of unknown length
* 15887 17089: contig of 1203 bp in length
* gap of unknown length
```



```
OY 61 cccgatttcttttttctcagaccgctgtcagtcgttcctgtaagaagcgatgat 120
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174321 CCCGATTTC-PTTGTGCTCTCAGACCGCTGTCAGTCGTCCTCCCGTAAGAAGCGATGAT- 174378

OY 121 ggcgtcgtggcgaattacacgcgcgtgaatggtttggcgtggattgcgtgcttgcctg 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174379 GCGCTGCTGGGCAATPACCTGCGGGCTAATGTTTGGCTGGGATGCGCT-CTTGGCCTG 174437

OY 181 cattgattatcgaaataagcctgctgcatacgcgtgattatggcggtggcctg 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174438 CAITTGATTATCGAAAAATGCCCTGGCTGCATACGCTGA-TATGGTGGCGGTGGCCTG 174496

OY 241 tatctctgctggatgggttacacatgctacgtggtgcactgaaagagcggtttct 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174497 TATCTCTGCTGGATGGTTTACCAGATGCTA-GTGGTGCACTGAAAAAGAGCGGTTTCT 174555

OY 301 gcacctggccacagtgctgaactggcgaaaaagtggcgagtttctcgaagtttactg 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174556 GCACCTGGCCACAGGTGGA-CTGGCGAAAAAGTGGCGCAGTTTCTTGAAAGGTTTACTG 174614

OY 361 accaatctcgctaaatccgaaagcgtattctactttgctcggtgttctcattgtttgtc 420
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174615 ACCAATCTCG-TAATCCGAAAGCATATCTACTTTGCTCGGTGTTCTCATTTGTTGTC 174673

OY 421 ggtgataacgttgccactaccgcgcgtggggcaatttttgcgctgcatctgtcgaaacg 480
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174674 -GTGATAACGTGGGCATPACCGCGCGCTGGGGCATTTTTCGCTGATCA-TGTCGAAACG 174731

OY 481 ctggcgtggtttaccgtgtgtccagcctggtttgccctgcgcgaatgcgcctggttat 540
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174732 CTGGCGTGGTTTACGCTGTTGCCAGCTGTTTGCCCTG-CGCAAAATGCGCGGTGTTAT 174790

OY 541 caactctggcgaaagtgatgatggttttgcggggcggttattgcgcatattgcatt 600
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174791 CAACGCTCGGCAAGTGGATGATGGTGTTC-CCGGGGCGTTATTTCGCGGATTGCGATT 174849

OY 601 catttgattatttcqcg 618
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 174850 CATTTGATTATTTCGCG 174867
```

STYSTD1	96086 bp	DNA	BCT	11-FEB-2000
LOCUS	Salmonella typhimurium fragment STMD1.			
DEFINITION	AF233324			
ACCESSION	AF233324.1	GI:6960215		
VERSION				
KEYWORDS	Salmonella typhimurium LT2.			
SOURCE	Salmonella typhimurium LT2			
ORGANISM	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
REFERENCE	1 (bases 1 to 96086)			
AUTHORS	Washington University Genome Sequencing Center.			
TITLE	The Salmonella typhimurium Genome Sequencing Project			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 96086)			
AUTHORS	Waterston,R.			
TITLE	Direct Submission			
JOURNAL	Submitted (09-FEB-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA			
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA e-mail: sclifton@watson.wustl.edu or jsplie@watson.wustl.edu			

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by
sequence from more than one m13 subclone.

NOTES:

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs
GLIMMER (Salzberg, S. Delcher, A., Kasif, A. and White, O. (1998)
NAR 26,544-548), and GeneMark (Lukashin, A.V. and Borodovsky, M.
(1998), NAR 26,1107-1115.

FEATURES	source
1. .96086	Location/Qualifiers
/organism="Salmonella typhimurium LT2"	
/strain="SGSC1412"	
/db_xref="taxon:99287"	
/clone="STMD1"	
<1. .980	
/note="rrfC"	
/product="23S ribosomal RNA"	
1072. .1191	
/note="rrfC"	
/product="5S ribosomal RNA"	
1244. .1323	
/product="tRNA-Asp"	
1329. .1401	
/product="tRNA-Trp"	
complement(1506. .2354)	
/gene="STMD1.99"	
complement(1506. .2354)	
/gene="STMD1.99"	
/note="83% identity with E. coli hypothetical protein (YifA) (SP:P22788) and 95% identity with amino acids 1-80 of E. coli possible regulatory protein (PSSR) (SP:P27826),"	
/codon_start=1	
/transl_table=11	
/evidence=not_experimental	
/protein_id="AAF33488.1"	
/db_xref="GI:6960298"	
/translation="MESTVDTLLKLEFVSTRHFGRAAALYLTOASVFRIROLE NQLGVLFRHNRNIRLTAGKLLPYAETLNTWQARKEVAHTSRINERSIGASAS LWECMLNMLGRILYQLOEPQSGQFPAIAQOSLVKOLHQRQLDLTLITAPKMDPE SSQLLGHFTIALYCSPARKKSELNLYRLEWGPDPFOHETGLIAADEVPVLTTSAEI ARQLSALNGCSWLVPVWANEKGLTVDATSATLSRPLYAIWLQNSDKYSILCDLLKT DVLDEQ"	
2461. .2799	
/gene="yifE"	
2461. .2799	
/gene="yifE"	
/note="STMD1.96"	
/codon_start=1	
/transl_table=11	
/product="95% identity with E. coli hypothetical protein (YifE) (SP:P27827)"	
/protein_id="AAF33485.1"	
/db_xref="GI:6960295"	
/translation="MAESFTTNRYPDNKHYPGRFSRHGDFTIKEAQLLRHGHAFND LDLGKRPVTEEEKLFVAVCGEREPTDAERVWSKYMTRIKPKRFTLSGGKQVQE GAEDYTEADD"	
complement(2824. .4344)	
/gene="yifB"	
complement(2824. .4344)	
/gene="yifB"	
/note="STMD1.95"	
/codon_start=1	
/transl_table=11	
/product="85% identity to E. coli hypothetical protein PF01078 (Magnesium chelatase, subunit ChlI), score=351.8, E=7.6e-102, N=1"	
/protein_id="AAF33484.1"	
/db_xref="GI:6960294"	
/translation="MSLAIVHTRAALGVNAPPTITIEVHISNGLPGLTMVGLPETTVKE	

RESULT 8
AC020885/c
LOCUS AC020885 268294 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RP23-46411, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020885
VERSION AC020885.2 GI:6980212
KEYWORDS HTG; HTGS.PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 268294)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 268294)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 16, 2000 this sequence version replaced gi:6686423.
* NOTE: This record contains 183 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 1013: contig of 1013 bp in length
* gap of unknown length
* 1014 1615: contig of 602 bp in length
* gap of unknown length
* 1616 2405: contig of 790 bp in length
* gap of unknown length
* 2406 3234: contig of 829 bp in length
* gap of unknown length
* 3235 3533: contig of 299 bp in length
* gap of unknown length
* 3534 4467: contig of 934 bp in length
* gap of unknown length
* 4468 5401: contig of 934 bp in length
* gap of unknown length
* 5402 6154: contig of 753 bp in length
* gap of unknown length
* 6155 6629: contig of 475 bp in length
* gap of unknown length
* 6630 7176: contig of 547 bp in length
* gap of unknown length
* 7177 7824: contig of 648 bp in length
* gap of unknown length
* 7825 8656: contig of 832 bp in length
* gap of unknown length
* 8657 8865: contig of 209 bp in length
* gap of unknown length
* 8866 9544: contig of 679 bp in length
* gap of unknown length
* 9545 10280: contig of 736 bp in length
* gap of unknown length
* 10281 10975: contig of 695 bp in length
* gap of unknown length
* 10976 11742: contig of 767 bp in length
* gap of unknown length
* 11743 12178: contig of 436 bp in length
* gap of unknown length
* 12179 12887: contig of 709 bp in length
* gap of unknown length
* 12888 13607: contig of 720 bp in length
* gap of unknown length
* 13608 13911: contig of 304 bp in length
* gap of unknown length
* 13912 14819: contig of 908 bp in length
* gap of unknown length
* 14820 15571: contig of 752 bp in length
* gap of unknown length
* 15572 15872: contig of 301 bp in length
* gap of unknown length
* 15873 16608: contig of 736 bp in length
* gap of unknown length
* 16609 16851: contig of 243 bp in length
* gap of unknown length
* 16852 17277: contig of 426 bp in length
* gap of unknown length
* 17278 18057: contig of 780 bp in length
* gap of unknown length
* 18058 18704: contig of 647 bp in length
* gap of unknown length
* 18705 19414: contig of 710 bp in length
* gap of unknown length
* 19415 20372: contig of 958 bp in length
* gap of unknown length
* 20373 21010: contig of 638 bp in length
* gap of unknown length
* 21011 21258: contig of 248 bp in length
* gap of unknown length
* 21259 22363: contig of 1105 bp in length
* gap of unknown length
* 22364 23639: contig of 1276 bp in length
* gap of unknown length
* 23640 24853: contig of 1214 bp in length
* gap of unknown length
* 24854 25795: contig of 942 bp in length
* gap of unknown length
* 25796 27027: contig of 1232 bp in length
* gap of unknown length
* 27028 27729: contig of 702 bp in length
* gap of unknown length
* 27730 28722: contig of 993 bp in length
* gap of unknown length
* 28723 29714: contig of 992 bp in length
* gap of unknown length
* 29715 30858: contig of 1144 bp in length
* gap of unknown length
* 30859 31397: contig of 539 bp in length
* gap of unknown length
* 31398 31671: contig of 274 bp in length
* gap of unknown length
* 31672 32959: contig of 1288 bp in length
* gap of unknown length
* 32960 33791: contig of 832 bp in length
* gap of unknown length
* 33792 34848: contig of 1057 bp in length
* gap of unknown length
* 34849 35060: contig of 212 bp in length
* gap of unknown length
* 35061 35776: contig of 716 bp in length
* gap of unknown length
* 35777 35903: contig of 127 bp in length
* gap of unknown length
* 35904 36602: contig of 699 bp in length
* gap of unknown length
* 36603 37110: contig of 508 bp in length
* gap of unknown length
* 37111 38059: contig of 949 bp in length
* gap of unknown length
* 38060 38730: contig of 671 bp in length
* gap of unknown length
* 38731 39790: contig of 1060 bp in length
* gap of unknown length
* 39791 40327: contig of 537 bp in length
* gap of unknown length
* 40328 40442: contig of 115 bp in length
* gap of unknown length

Mon May 7 11:08:10 2001

		gap of unknown length	
*	40443	contig of 1246 bp in length	
*		gap of unknown length	
*	41689	contig of 1000 bp in length	
*		gap of unknown length	
*	42689	contig of 1159 bp in length	
*		gap of unknown length	
*	43848	contig of 664 bp in length	
*		gap of unknown length	
*	44512	contig of 1269 bp in length	
*		gap of unknown length	
*	45781	contig of 1422 bp in length	
*		gap of unknown length	
*	47203	contig of 1445 bp in length	
*		gap of unknown length	
*	48648	contig of 1005 bp in length	
*		gap of unknown length	
*	49653	contig of 833 bp in length	
*		gap of unknown length	
*	50486	contig of 1042 bp in length	
*		gap of unknown length	
*	51528	contig of 1592 bp in length	
*		gap of unknown length	
*	53120	contig of 504 bp in length	
*		gap of unknown length	
*	53624	contig of 1046 bp in length	
*		gap of unknown length	
*	54670	contig of 640 bp in length	
*		gap of unknown length	
*	55310	contig of 1073 bp in length	
*		gap of unknown length	
*	56383	contig of 243 bp in length	
*		gap of unknown length	
*	56626	contig of 720 bp in length	
*		gap of unknown length	
*	57346	contig of 160 bp in length	
*		gap of unknown length	
*	57506	contig of 1403 bp in length	
*		gap of unknown length	
*	58909	contig of 1060 bp in length	
*		gap of unknown length	
*	59969	contig of 1409 bp in length	
*		gap of unknown length	
*	61378	contig of 914 bp in length	
*		gap of unknown length	
*	62292	contig of 1028 bp in length	
*		gap of unknown length	
*	63320	contig of 499 bp in length	
*		gap of unknown length	
*	63819	contig of 1005 bp in length	
*		gap of unknown length	
*	64824	contig of 446 bp in length	
*		gap of unknown length	
*	65270	contig of 635 bp in length	
*		gap of unknown length	
*	65905	contig of 911 bp in length	
*		gap of unknown length	
Query Match 67.3%; Score 415.8; DB 66; Length 268294;			
Best Local Similarity 99.5%; Pred. No. 3.8e-111;			
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	200	tggcctgctgcacgcgtgattatggcggtggcggtggcctgtatctctgctgagtggtt 259	
Db	11742	TGCCCTTGGCTGCATACGCTGATATATGGTGGCGGTGGCCCTGTATCTCTCGATGGGTT 11683	
Qy	260	accagatgctagctggtgcactgaaaaagagcggttttctgcacctgcgcacaggttcg 319	
Db	11682	ACCAATGCTAGCTGCTGCACACAAAAGAGCGGTTTCTGCACCTGCGCCACAGGTGCG 11623	
Qy	320	agctgcaaaaagtggcgagcttctctgaaagggtttactgaccaatctcgctaataccga 379	
Db	11622	AGCTGGCGAAAAGTGGCGCAGTTTCTCTGAAAAGTTTACTGACCAATCTCGCTAATCCGA 11563	

QY	380	aagcgattatctacttggctgctggtgttctcattgttttctcggtagataacgttggcacta 439	
Db	11562	AAGCGATTATCTACTTTGGCTCGGTGTTCATATGTTTGTTCGGTGATACAGTTGGCACCTA 11503	
QY	440	ccscgcctggggcatttttggcctgctgctcattgcgaacgcgtggcgtgtttaccgtcg 499	
Db	11502	CCCGCGCTGGGGCANTTTTGGCTGCTATCATTTCTCGAAAGCGTGGCTGTTTACCGTCG 11443	
QY	500	ttgccagcctgttttgcctgcgcgcaaatgcgcgtgggttatcaacgtctgcggaagtgga 559	
Db	11442	TTGCCAGCCTGTTTGGCCCTGCGCAATGCGCGTGGTGTATCAACGCTCTGCGCAAGTGGGA 11383	
QY	560	ttgatggtttgcggggcgatttgcggatttgcgcatttgcatttgcatttgcgg 618	
Db	11382	TTGATGGTTTGGCGGGCGGCTTATTTCGCCGATTTGGCATTTCATTGATTATTTCGCGG 11324	

RESULT 9

AC079167 290452 bp DNA HTG 30-AUG-2000

LOCUS Mus musculus chromosome 6 clone RP23-226023 strain C57BL6/J, ***

DEFINITION SEQUENCING IN PROGRESS ***, 196 unordered pieces.

AC079167

VERSION AC079167.2 GI:9945047

KEYWORDS HTG; HTGS PHASE1.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 290452)

AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE High Throughput Mouse Sequencing

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 290452)

AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE Direct Submission

JOURNAL Submitted (23-AUG-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

COMMENT On Aug 30, 2000 this sequence version replaced gi:9885883.

-----Genome Center Center: Albert Einstein College of Medicine Center

Code: AECOM

Web site: http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSeq/mouseSeqtable.hts

Contact: jhan@sequence.aecom.yu.edu

-----Summary Statistics

Center project name: AEN

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

*Consensus quality: 196511 at least Q20

*Consensus quality: 172308 at least Q30

*Consensus quality: 138790 at least Q40

Estimated insert size: agarose-FP - N/A

**Estimated insert size: 286552 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 3.3x sum-of-contigs - N/A

* NOTE: This is a 'working draft' sequence. It currently consists of 196 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

* 1 852: contig of 852 bp in length

*

```
*      853      872: gap of unknown length
*      873      1753: contig of 881 bp in length
*     1754      1773: gap of unknown length
*     1774      2658: contig of 885 bp in length
*     2659      2678: gap of unknown length
*     2679      3514: contig of 836 bp in length
*     3515      3534: gap of unknown length
*     3535      4449: contig of 915 bp in length
*     4450      4469: gap of unknown length
*     4470      5384: contig of 915 bp in length
*     5385      5404: gap of unknown length
*     5405      6273: contig of 869 bp in length
*     6274      6293: gap of unknown length
*     6294      7095: contig of 802 bp in length
*     7096      7115: gap of unknown length
*     7116      7981: contig of 866 bp in length
*     7982      8001: gap of unknown length
*     8002      8729: contig of 728 bp in length
*     8730      8749: gap of unknown length
*     8750      9879: contig of 1126 bp in length
*     9880      9899: gap of unknown length
*     9896      10631: contig of 736 bp in length
*     10632      10651: gap of unknown length
*     10652      11577: contig of 926 bp in length
*     11578      11597: gap of unknown length
*     11598      12454: contig of 857 bp in length
*     12455      12474: gap of unknown length
*     12475      13334: contig of 860 bp in length
*     13335      13354: gap of unknown length
*     13355      14279: contig of 925 bp in length
*     14280      14299: gap of unknown length
*     14300      15245: contig of 946 bp in length
*     15246      15265: gap of unknown length
*     15266      16058: contig of 793 bp in length
*     16059      16079: gap of unknown length
*     16079      16946: contig of 867 bp in length
*     16946      16965: gap of unknown length
*     16966      17850: contig of 885 bp in length
*     17851      17870: gap of unknown length
*     17871      18722: contig of 852 bp in length
*     18723      18743: gap of unknown length
*     18743      19557: contig of 814 bp in length
*     19557      19577: gap of unknown length
*     19577      20816: contig of 1240 bp in length
*     20817      20836: gap of unknown length
*     20837      21756: contig of 920 bp in length
*     21757      21776: gap of unknown length
*     21777      22511: contig of 741 bp in length
*     22518      22537: gap of unknown length
*     22538      23457: contig of 920 bp in length
*     23458      23477: gap of unknown length
*     23478      24289: contig of 812 bp in length
*     24290      24309: gap of unknown length
*     24310      25249: contig of 940 bp in length
*     25250      25269: gap of unknown length
*     25270      26136: contig of 866 bp in length
*     26136      27048: contig of 893 bp in length
*     27049      27068: gap of unknown length
*     27069      28003: contig of 935 bp in length
*     28004      28023: gap of unknown length
*     28024      28874: contig of 851 bp in length
*     28875      28894: gap of unknown length
*     28895      29777: contig of 883 bp in length
*     29778      30160: contig of 363 bp in length
*     30161      30180: gap of unknown length
*     30181      31014: contig of 834 bp in length
*     31015      31034: gap of unknown length
*     31035      31880: contig of 846 bp in length
*     31881      31900: gap of unknown length
*     31901      33160: contig of 1260 bp in length
*     33161      33180: gap of unknown length

33181      34033: contig of 853 bp in length
*     34034      34053: gap of unknown length
*     34054      34972: contig of 919 bp in length
*     34973      34992: gap of unknown length
*     34993      35875: contig of 883 bp in length
*     35876      35895: gap of unknown length
*     35896      36731: contig of 836 bp in length
*     36732      36751: gap of unknown length
*     36752      37639: contig of 888 bp in length
*     37640      37659: gap of unknown length
*     37660      38499: contig of 840 bp in length
*     38500      38519: gap of unknown length
*     38520      39348: contig of 829 bp in length
*     39349      39368: gap of unknown length
*     39369      40377: contig of 1009 bp in length
*     40378      40397: gap of unknown length
*     40398      41469: contig of 1072 bp in length
*     41470      41489: gap of unknown length
*     41490      41890: contig of 401 bp in length
*     41891      41910: gap of unknown length
*     41911      42795: contig of 885 bp in length
*     42796      42815: gap of unknown length
*     42816      43744: contig of 929 bp in length
*     43745      43764: gap of unknown length
*     43765      44628: contig of 864 bp in length
*     44629      44648: gap of unknown length
*     44649      45958: contig of 1310 bp in length
*     45959      45978: gap of unknown length
*     45979      46827: contig of 849 bp in length
*     46828      46847: gap of unknown length
*     46848      47807: contig of 960 bp in length
*     47808      47827: gap of unknown length
*     47828      48376: contig of 549 bp in length
*     48377      48396: gap of unknown length
*     48397      49297: contig of 901 bp in length
*     49298      49317: gap of unknown length
*     49318      50182: contig of 865 bp in length
*     50183      50202: gap of unknown length
*     50203      51054: contig of 852 bp in length
*     51055      51074: gap of unknown length
*     51075      51373: contig of 299 bp in length
*     51374      51393: gap of unknown length
*     51394      52053: contig of 660 bp in length
*     52054      52073: gap of unknown length
*     52074      52944: contig of 871 bp in length
*     52945      52964: gap of unknown length
*     52965      53539: contig of 575 bp in length
*     53540      53559: gap of unknown length
*     53560      54374: contig of 815 bp in length
*     54375      54394: gap of unknown length
*     54395      55253: contig of 859 bp in length
*     55254      55273: gap of unknown length
*     55274      56527: contig of 1254 bp in length
*     56528      56547: gap of unknown length
*     56548      57376: contig of 829 bp in length
*     57377      57396: gap of unknown length
*     57397      57473: contig of 77 bp in length
*     57474      57493: gap of unknown length
*     57494      58337: contig of 844 bp in length
*     58338      58357: gap of unknown length
*     58358      58444: contig of 87 bp in length
*     58445      58464: gap of unknown length
*     58465      59288: contig of 824 bp in length
*     59289      59308: gap of unknown length
*     59309      60499: contig of 1191 bp in length
*     60500      60519: gap of unknown length
*     60520      61859: contig of 1340 bp in length
*     61860      61879: gap of unknown length
*     61880      62752: contig of 873 bp in length
*     62753      62772: gap of unknown length
*     62773      63623: contig of 851 bp in length
*     63624      63643: gap of unknown length
*     63644      64694: contig of 1051 bp in length
```



```

* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
* 1
* 296: contig of 296 bp in length
* 297 gap of unknown length
* 1049: contig of 753 bp in length
* 1050 gap of unknown length
* 1812: contig of 763 bp in length
* 1813 gap of unknown length
* 2420: contig of 608 bp in length
* 2421 gap of unknown length
* 2953: contig of 533 bp in length
* 2954 gap of unknown length
* 3645: contig of 692 bp in length
* 3646 gap of unknown length
* 4766: contig of 1121 bp in length
* 4767 gap of unknown length
* 5296: contig of 530 bp in length
* 5297 gap of unknown length
* 5584: contig of 288 bp in length
* 5585 gap of unknown length
* 6454: contig of 870 bp in length
* 6455 gap of unknown length
* 6853: contig of 399 bp in length
* 6854 gap of unknown length
* 7566: contig of 713 bp in length
* 7567 gap of unknown length
* 8253: contig of 687 bp in length
* 8254 gap of unknown length
* 9058: contig of 805 bp in length
* 9059 gap of unknown length
* 9678: contig of 620 bp in length
* 9679 gap of unknown length
* 10272: contig of 594 bp in length
* 10273 gap of unknown length
* 10890: contig of 618 bp in length
* 10891 gap of unknown length
* 11695: contig of 805 bp in length
* 11696 gap of unknown length
* 12677: contig of 982 bp in length
* 12678 gap of unknown length
* 13130: contig of 453 bp in length
* 13131 gap of unknown length
* 13878: contig of 748 bp in length
* 13879 gap of unknown length
* 14026: contig of 148 bp in length
* 14027 gap of unknown length
* 14869: contig of 843 bp in length
* 14870 gap of unknown length
* 15738: contig of 869 bp in length
* 15739 gap of unknown length
* 16426: contig of 888 bp in length
* 16427 gap of unknown length
* 17161: contig of 735 bp in length
* 17162 gap of unknown length
* 17938: contig of 777 bp in length
* 17939 gap of unknown length
* 18715: contig of 777 bp in length
* 18716 gap of unknown length
* 19493: contig of 778 bp in length
* 19494 gap of unknown length
* 20212: contig of 719 bp in length
* 20213 gap of unknown length
* 20780: contig of 568 bp in length
* 20781 gap of unknown length
* 21499: contig of 719 bp in length
* 21500 gap of unknown length
* 21570: contig of 71 bp in length
* gap of unknown length
*
* 21571 contig of 808 bp in length
* 22378: gap of unknown length
* 22379 contig of 562 bp in length
* 22941 gap of unknown length
* 23943: contig of 1003 bp in length
* 23944 gap of unknown length
* 24650: contig of 707 bp in length
* 25686: gap of unknown length
* 25687 contig of 1036 bp in length
* 25820: gap of unknown length
* 25821 contig of 134 bp in length
* 26268: gap of unknown length
* 26269 contig of 448 bp in length
* 27152: gap of unknown length
* 27153 contig of 884 bp in length
* 28057: gap of unknown length
* 28058 contig of 905 bp in length
* 29002: gap of unknown length
* 29003 contig of 945 bp in length
* 29991: gap of unknown length
* 29992 contig of 989 bp in length
* 30780: gap of unknown length
* 30781 contig of 789 bp in length
* 31514: gap of unknown length
* 31515 contig of 734 bp in length
* 32152: gap of unknown length
* 32153 contig of 638 bp in length
* 32933: gap of unknown length
* 32934 contig of 781 bp in length
* 33042: gap of unknown length
* 33043 contig of 109 bp in length
* 34310: gap of unknown length
* 34311 contig of 1268 bp in length
* 34992: gap of unknown length
* 34993 contig of 682 bp in length
* 35821: gap of unknown length
* 35822 contig of 829 bp in length
* 37013: gap of unknown length
* 37014 contig of 1192 bp in length
* 38147: gap of unknown length
* 38148 contig of 1134 bp in length
* 38852: gap of unknown length
* 38853 contig of 705 bp in length
* 39840: gap of unknown length
* 39841 contig of 988 bp in length
* 40874: gap of unknown length
* 40875 contig of 1034 bp in length
* 42387: gap of unknown length
* 42388 contig of 1513 bp in length
* 43258: gap of unknown length
* 43259 contig of 871 bp in length
* 44365: gap of unknown length
* 44366 contig of 1107 bp in length
* 45674: gap of unknown length
* 45675 contig of 1309 bp in length
* 47108: gap of unknown length
* 47109 contig of 1434 bp in length
* 48115: gap of unknown length
* 48116 contig of 1007 bp in length
* 49871: gap of unknown length
* 49872 contig of 1756 bp in length
* 51661: gap of unknown length
* 51662 contig of 1790 bp in length
* 53728: gap of unknown length
* 53729 contig of 2067 bp in length
* 56784: gap of unknown length
* 56785 contig of 3056 bp in length
* 59220: gap of unknown length
* 59221 contig of 2436 bp in length
* 62274: gap of unknown length
* 62275 contig of 3054 bp in length.
*
* Location/Qualifiers
* 1. 62274

```

FEATURES
source

* 43205 44325: contig of 1121 bp in length
* 44326 44425: gap of unknown length
* 44426 45890: contig of 1465 bp in length
* 45891 45990: gap of unknown length
* 45991 47229: contig of 1239 bp in length
* 47230 47329: gap of unknown length
* 47330 48645: contig of 1316 bp in length
* 48646 48745: gap of unknown length
* 48746 49980: contig of 1235 bp in length
* 49981 50080: gap of unknown length
* 50081 51209: contig of 1129 bp in length
* 51210 51309: gap of unknown length
* 51310 52798: contig of 1489 bp in length
* 52799 52898: gap of unknown length
* 52899 54154: contig of 1256 bp in length
* 54155 54254: gap of unknown length
* 54255 55552: contig of 1298 bp in length
* 55553 55652: gap of unknown length
* 55653 57780: contig of 2128 bp in length
* 57781 57880: gap of unknown length
* 57881 59692: contig of 1812 bp in length
* 59693 59792: gap of unknown length
* 59793 60947: contig of 1155 bp in length
* 60948 61047: gap of unknown length
* 61048 62132: contig of 1085 bp in length
* 62133 62332: gap of unknown length
* 62333 63505: contig of 1273 bp in length
* 63506 63605: gap of unknown length
* 63606 65114: contig of 1509 bp in length
* 65115 65215: gap of unknown length
* 65215 67705: contig of 2491 bp in length
* 67706 67805: gap of unknown length
* 67806 69266: contig of 1461 bp in length
* 69267 69366: gap of unknown length
* 69367 70408: contig of 1042 bp in length
* 70409 70508: gap of unknown length
* 70509 72169: contig of 1661 bp in length
* 72170 72269: gap of unknown length
* 72270 73711: contig of 1442 bp in length
* 73712 73811: gap of unknown length
* 73812 76440: contig of 2629 bp in length
* 76441 76540: gap of unknown length
* 76541 77916: contig of 1376 bp in length
* 77917 78016: gap of unknown length
* 78017 80190: contig of 2174 bp in length
* 80191 80290: gap of unknown length
* 80291 81379: contig of 1089 bp in length
* 81380 81479: gap of unknown length
* 81480 82828: contig of 1349 bp in length
* 82829 82928: gap of unknown length
* 82929 86817: contig of 3889 bp in length
* 86818 86917: gap of unknown length
* 86919 89152: contig of 2235 bp in length
* 89153 89252: gap of unknown length
* 89253 91565: contig of 2313 bp in length
* 91566 91665: gap of unknown length
* 91666 94073: contig of 2408 bp in length
* 94074 94173: gap of unknown length
* 94174 98664: contig of 4491 bp in length
* 98665 98764: gap of unknown length
* 98765 101607: contig of 2843 bp in length
* 101608 101707: gap of unknown length
* 101708 103901: contig of 2194 bp in length
* 103902 104001: gap of unknown length
* 104002 105900: contig of 1899 bp in length
* 105901 106000: gap of unknown length
* 106001 109702: contig of 3702 bp in length
* 109703 109802: gap of unknown length
* 109803 116164: contig of 6362 bp in length
* 116165 116264: gap of unknown length
* 116265 126230: contig of 9966 bp in length
* 126231 126330: gap of unknown length
* 126331 145648: contig of 19318 bp in length

* 145649 145748: gap of unknown length
* 145749 164949: contig of 19201 bp in length
* 164950 165049: gap of unknown length
* 165050 198403: contig of 33354 bp in length
* 198404 198503: gap of unknown length
* 198504 243378: contig of 44875 bp in length
* 243379 298166: contig of 54688 bp in length.
* 243479 298166: contig of 54688 bp in length.
FEATURES
source
1..298166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-358F6"
/clone_lib="RP11 human BAC library 11"
Query Match 21.5%; Score 132.8; DB 78; Length 298166;
Best Local Similarity 60.9%; Pred. No. 5.6e-28;
Matches 266; Conservative 0; Mismatches 168; Indels 3; Gaps 3;
QY 4 ttgatgtatttctcacccgtcgccatggtgcacattgtggcgcttatgagcccggtccc 63
Db 90711 TTAGAGGTATTTTACACTAANAATCGTCACATTGTGGCGCTGATGA-CCAAGGGCCC 90769
QY 64 gattcttttttgcctcagacacgctgcagtcgtcccgtaaaagagcgatgatgggc 123
Db 90770 TATTTCTTATGGGATGTGAGACCGGAGTCATTCTTCGGGAAGAGAATCTACGATGGG 90829
QY 124 gtctgtggcattaccctgcggcgtaattgttggcgctggattgcgtgttggcctgc 183
Db 90830 GAGGTGGCCATTACCTCGTCTGATGGTTTGGCTGGGATAGCTCTGCTCAGCGCAT 90889
QY 184 ttgattatcgaaaaaatggcctgctgcatacgtgattatgttggcggtggtcctgtat 243
Db 90890 GGGG-TATTCAAAGAAATGATCTGTGTATACTCCTGTAGCCGCGC-GTTGCCGGAA 90947
QY 244 ctctgtggatgggttaccagatgctacgtgtgctacgtgaaagagggcggtttctgca 303
Db 90948 CGATGTTGAAAGAGTTAAAGATGCTTGTGTCGCGCGTGATGGAGGAGACGTCCTCTTC 91007
QY 304 cctgcgcacaggtcgagctggcgaaaagtgggcgagtttccctgaaggtttactgacc 363
Db 91008 TCTGTGTCTAGGGGGAGGTGCTTAACTAGGGGGGCAACTCGGTGAGAGGGCCACTGACC 91067
QY 364 aatctcgctaatccgaaagcattatctaatttggcgtggttctcattgttctcggt 423
Db 91068 ATCACCCCATCCCTGAAACTATTAAAGCCGTAGAGTGGCGATCCGAACTCGGTGGGT 91127
QY 424 gataacgttggcactac 440
Db 91128 GTTAATGGGAATCCAC 91144

RESULT 13
AC022157 AC022157 4833 bp DNA HTG 26-JAN-2000
LOCUS Homo sapiens chromosome 5 clone RP1-280K18, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC022157
VERSION AC022157.1 GI:6758618
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4833)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4833)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint

COMMENT	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA			Unpublished		
	-----Genome Center			2 (bases 1 to 265383)		
	Center: Joint Genome Institute			DOE Joint Genome Institute.		
	Center Code: JGI			Direct Submission		
	Web site: http://www.jgi.doe.gov			Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
FEATURES	*****			On Feb 16, 2000 this sequence version replaced gi:6686434.		
	* NOTE: This record contains 8 individual			* sequencing reads that have not been assembled into		
	* contigs. Runs of N are used to separate the reads			* and the order in which they appear is completely		
	* arbitrary. Low-pass sequence sampling is useful for			* identifying clones that may be gene-rich and allows		
	* overlap relationships among clones to be deduced.			* However, it should not be assumed that this clone		
source	* will be sequenced to completion. In the event that			* the record is updated, the accession number will		
	* be preserved.			* be preserved.		
	* 1			869: contig of 869 bp in length		
	* 837			gap of unknown length		
	* 1567			contig of 730 bp in length		
LOCUS	* 2613			gap of unknown length		
	* 2722			contig of 1046 bp in length		
	* 2830			gap of unknown length		
	* 3053			contig of 109 bp in length		
	* 4257			gap of unknown length		
DEFINITION	* 4257			contig of 108 bp in length		
	* 4257			gap of unknown length		
	* 4257			contig of 223 bp in length		
	* 4257			gap of unknown length		
	* 4257			contig of 1204 bp in length		
ACCESSION	* 4257			gap of unknown length		
	* 4257			contig of 577 bp in length.		
	* 4257			Location/Qualifiers		
	* 4257			1. .4833		
	* 4257			/organism="Homo sapiens"		
VERSION	* 4257			/db_xref="taxon:9606"		
	* 4257			/chromosome="5"		
	* 4257			/clone="RP1-280K18"		
	* 4257			BASE COUNT 1218 a 1298 c 1250 g 1055 t 12 others		
	* 4257			ORIGIN		
KEYWORDS	* 4257			Query Match 19.4%; Score 120; DB 67; Length 4833;		
	* 4257			Best Local Similarity 91.4%; Pred. No. 2.4e-24;		
	* 4257			Matches 139; Conservative 0; Mismatches 10; Indels 3; Gaps 1;		
	* 4257			Qy 449 gggcgatttttgcgtgatcattgtcgaaacgtgcgtggtttaccgtcg---ttgccca 505		
	* 4257			Db 25 GCGCATTTTATAGCTGATCATTTGTGAAACGCTGGCGTGGTTACCGCTCGTTGCCAC 84		
SOURCE	* 4257			Qy 506 gctgtttgcccctgcgcaaatgcgcgtggttatcaacgtctgcggaagtgtgattgatg 565		
	* 4257			Db 85 GCCTGTTTGGCCCTGCGCAAAATGCGCCGCGGTATACACGCTCGCGGAAGTGATTGATG 144		
	* 4257			Qy 566 gttttcgccggcggttatttgcggatttggc 597		
	* 4257			Db 145 GTTTTGCCGGGGCGGTATTATTTGCCGATTTTGGC 176		
	* 4257			RESULT 14		
ORGANISM	* 4257			AC020874 265383 bp DNA HTG 16-FEB-2000		
	* 4257			LOCUS AC020874/c		
	* 4257			DEFINITION Mus musculus clone RP23-333M14, LOW-PASS SEQUENCE SAMPLING.		
	* 4257			ACCESSION AC020874		
	* 4257			VERSION AC020874.2 GI:6980207		
REFERENCE	* 4257			KEYWORDS HTG; HTGS_PHASE0.		
	* 4257			SOURCE house mouse.		
	* 4257			ORGANISM Mus musculus		
	* 4257			Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	* 4257			Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	* 4257			1 (bases 1 to 265383)		
	* 4257			DOE Joint Genome Institute.		
	* 4257			Sequencing of Mouse		
	* 4257			TITLE		
	* 4257			Sequencing of Mouse		

```
*
* 19194 gap of unknown length
* 19737: contig of 544 bp in length
* 19738 gap of unknown length
* 20600: contig of 863 bp in length
* 21239: contig of 639 bp in length
* 21240 gap of unknown length
* 21903: contig of 664 bp in length
* 21904 gap of unknown length
* 22705: contig of 802 bp in length
* 22706 gap of unknown length
* 22948: contig of 243 bp in length
* 22949 gap of unknown length
* 23682: contig of 734 bp in length
* 23683 gap of unknown length
* 24364: contig of 682 bp in length
* 24365 gap of unknown length
* 25288: contig of 924 bp in length
* 25289 gap of unknown length
* 25954: contig of 666 bp in length
* 25955 gap of unknown length
* 26094: contig of 140 bp in length
* 26095 gap of unknown length
* 27015: contig of 921 bp in length
* 27016 gap of unknown length
* 27740: contig of 725 bp in length
* 27741 gap of unknown length
* 28060: contig of 320 bp in length
* 28061 gap of unknown length
* 28343: contig of 283 bp in length
* 28344 gap of unknown length
* 29016: contig of 673 bp in length
* 29017 gap of unknown length
* 29371: contig of 355 bp in length
* 29372 gap of unknown length
* 29586: contig of 215 bp in length
* 29587 gap of unknown length
* 29691: contig of 105 bp in length
* 29692 gap of unknown length
* 30501: contig of 810 bp in length
* 30502 gap of unknown length
* 31132: contig of 631 bp in length
* 31133 gap of unknown length
* 31249: contig of 117 bp in length
* 31250 gap of unknown length
* 32467: contig of 1218 bp in length
* 32468 gap of unknown length
* 32878: contig of 411 bp in length
* 32879 gap of unknown length
* 34069: contig of 1191 bp in length
* 34070 gap of unknown length
* 34694: contig of 625 bp in length
* 34695 gap of unknown length
* 35933: contig of 1239 bp in length
* 35934 gap of unknown length
* 36920: contig of 987 bp in length
* 36921 gap of unknown length
* 37173: contig of 253 bp in length
* 37174 gap of unknown length
* 37282: contig of 109 bp in length
* 37283 gap of unknown length
* 38196: contig of 914 bp in length
* 38197 gap of unknown length
* 38452: contig of 256 bp in length
* 38453 gap of unknown length
* 38678: contig of 226 bp in length
* 38679 gap of unknown length
* 38835: contig of 157 bp in length
* 38836 gap of unknown length
* 39561: contig of 726 bp in length
* 39562 gap of unknown length
* 40452: contig of 891 bp in length
* gap of unknown length
```

```
* 40453 41223: contig of 771 bp in length
* gap of unknown length
* 41224 41864: contig of 641 bp in length
* gap of unknown length
* 41865 42690: contig of 826 bp in length
* gap of unknown length
* 42691 42988: contig of 298 bp in length
* gap of unknown length
* 42989 43612: contig of 624 bp in length
* gap of unknown length
* 43613 44265: contig of 653 bp in length
* gap of unknown length
* 44266 45263: contig of 998 bp in length
* gap of unknown length
* 45264 46592: contig of 1329 bp in length
* gap of unknown length
* 46593 47054: contig of 462 bp in length
* gap of unknown length
* 47055 47193: contig of 139 bp in length
* gap of unknown length
* 47194 48027: contig of 834 bp in length
* gap of unknown length
* 48028 48670: contig of 643 bp in length
* gap of unknown length
* 48671 49511: contig of 841 bp in length
* gap of unknown length
* 49512 49895: contig of 384 bp in length
* gap of unknown length
* 49896 50124: contig of 229 bp in length
* gap of unknown length
* 50125 50776: contig of 652 bp in length
* gap of unknown length
* 50777 50905: contig of 129 bp in length
* gap of unknown length
* 50906 51260: contig of 355 bp in length
* gap of unknown length
* 51261 52029: contig of 769 bp in length
* gap of unknown length
* 52030 52294: contig of 265 bp in length
* gap of unknown length
* 52295 53140: contig of 846 bp in length
* gap of unknown length
*
Query Match 14.8%; Score 91.4; DB 66; Length 265383;
Best Local Similarity 90.1%; Pred. No. 8.2e-16;
Matches 109; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 1 atgttgatgtattctcaccgtgcacacgtgtgcacattgtgcgc-ttatgagccgg 59
Db 124944 ATGTTGATGTTATTCTACCGTCGCCATTTGTCGACATTGTCGCGCATTTATGAGCCCGG 124885
QY 60 tccgattcttttctctcagaccgtgcacgtctccgttaaagaagcgatgat 119
Db 124884 TCCCGATTTCTTTTGTCTCTCAGACCGCTGTCAGTCGTCCTCGTAAGAGATATTCT 124825
QY 120 g 120
Db 124824 G 124824
RESULT 15
AE006110/c 10184 bp DNA BCT 07-FEB-2001
LOCUS Pasteurella multocida PM70 section 77 of 204 of the complete
DEFINITION genome.
ACCESSION AE006110 AE004439
VERSION AE006110.1 GI:12721018
KEYWORDS Pasteurella multocida.
SOURCE Pasteurella multocida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
REFERENCE 1 (bases 1 to 10184)
```

AUTHORS May,B.J., Zhang,Q., Li,L., Paustian,M.L., Whittam,T.S. and Kapur,V.S.

TITLE Complete nucleotide sequence of an avian isolate of Pasteurella multocida

JOURNAL Proc. Natl. Acad. Sci. U.S.A. (2001) In press

REFERENCE 2 (bases 1 to 10184)

AUTHORS Zhang,Q. and Kapur,V.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-2000) Department of Veterinary Pathobiology,

University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN

55108, USA

FEATURES Location/Qualifiers

source i.10184

/organism="Pasteurella multocida"

/strain="PM70"

/db_xref="taxon:747"

105..1820

/gene="ttrS"

/note="PM0724"

105..1820

/gene="ttrS"

/codon_start=1

/transl_table=11

/product="ttrS"

/protein_id="AAK02808.1"

/db_xref="GI:12721019"

/translation="MKKCSFLKMFILLSEPMALVYAQWKIGILAQRGVAHTATW
OPWIDNLNQLSPHQFTLVTDLNDLIQDAQGVDFILSNQAQFFYLNNNEVRWLVL
RSPRIQSSMTGRIGSAILVRQSPFPQLDRLHKTVSAVGDNAFGGLGILYEFYQO
GLVENKDFOLKFTGFPVDNITLLIQERKVDAAIIVPGVILBELVNEGYYKKEDFRLLAG
KENVLGCLA STELLPNWSLAALSOVDPDLASKLVTLTLLSOVPAHLPOWTPYSTSQAD
TVLGLKHPQOQLWDSKYWLOQYCGWIIFVLLLVNLVWISFQIHRKSKALQQAAY
REMHYEQQLQADRLLSILGEMSTGIAHEINQPLSARMYVEGVKYHLAQLPAKNTIV
EILDPLQOQVDRSDIIRNLNNAWAKKKTPSSCEVALKPLQQRVYSFMSLQTHQKAQL
NLNCPASKIKIDITLLEQVLCNLLNASQAGSEIEVNVKAVGEYMTIHIVDNGSGF
KAGELAPFPVFRSSKANGLGGLVLCQRLMRISINGDISLHNRLDQOQGAEVILKLPNK
VGETC"

1814..2392

/gene="ttrR"

/note="PM0725"

1814..2392

/gene="ttrR"

/codon_start=1

/transl_table=11

/product="ttrR"

/protein_id="AAK02809.1"

/db_xref="GI:12721020"

/translation="MLIHLVDDDLTVLDAARFLLEQAGYQVOTWSNQAQFVNKAPLPE
PGVLLDMKMPLLDGHQVHFQCOQOSTLAYVIMTAHADVPMAVQELKQGAVDLQKP
VOFSOLOSLVLTATQTTQSYRYKIRHCYAQLSRKELDILALLIQGCINRQIAEALN
LSVTVVEHRSHIMEKMOAQTIAELIYKTAQL"

complement(2427..3239)

/gene="dapB"

/note="PM0726"

complement(2427..3239)

/gene="dapB"

/codon_start=1

/transl_table=11

/product="dapB"

/protein_id="AAK02810.1"

/db_xref="GI:12721021"

/translation="MTRVAIVGAGGRMRQLIQAVSETEGVVLGAAPERQSGSLLGA
DAGELANYHLGVQITDOLASQODFDLLIDFTRPEGLAHLAFCAVQAKNNKVIQTTG
FDDGKAKIQQAADSIGIVFASNFSGVNLVFKLEAKAQMVDYCDIEIIEBAHRHK
VDAPSGTALSNGEHAKTGLRDLKTHGVFTRDGLTGERKDEIGFATIRASDVGEHS
WMFADIGERIEAHKATSRWIFAKGAVRAAKWLAQKEGLFDMTVDLNLNQL"

complement(3294..3938)

/gene="PM0727"

complement(3294..3938)

/gene="PM0727"

/codon_start=1

/transl_table=11

/product="unknown"

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/protein_id="AAK02811.1"
/db_xref="GI:12721022"
/translation="MNLIIIVHFFGLITPGDFFVVSRLAASNSRNALCAVIGITL
GVLFALASILGLAILFNTVPVLOGLWMFLGGYLAJGLVLMKSOQNVFEPVSEOE
QNKQTSIKKEITLGNLNSNAKAVIYFASVMSLVLYNLTQTQWINSALLIIVLETFE
YFAISIVFSROAQKQFYQYSRYIDHLSGVIFLLFGLVYMLSGIGEMTFLISK"
complement(3944..4429)
/gene="pgpA"
/note="PM0728"
complement(3944..4429)
/gene="pgpA"
/codon_start=1
/transl_table=11
/product="PgpA"
/protein_id="AAK02812.1"
/db_xref="GI:12721023"
/translation="MLKTPLOQLKSNPITHFLALGFCAGLIRPAGPTWGSVMGLMLGW
GLLOYLNLGSGFFLTCALCFALGCYLCOKTAEADMGVHDHGAIVWDEIVGIFVLLSLPN
LSPFWCLTAFILFRFFDILKPYPIRYFDHKLSESGFGIMLDDIILAAIYAVLVILRLM
F"
complement(4439..5431)
/gene="thil"
/note="PM0729"
complement(4439..5431)
/gene="thil"
/codon_start=1
/transl_table=11
/product="Thil"
/protein_id="AAK02813.1"
/db_xref="GI:12721024"
/translation="MSHGEFDILQRYFTASKRTARKDVILSIGDCCAITELRONORLV
ITDTMVNTHFLPMSHAADLAYKAAVTNLSDLAAMGAEPANWSLATLFDVDAQWLS
TDSQSFDDLDHYNDVITGGDTTKGPSLVTITAHPIIPKGLGCLRSASVGDMDIYISG
TFGSAAGLSLLSQOTPLNSDHEFLIQRHLRPTPTVLLGLELAVSSANAADISDG
FLADLDHLTRSCQGVTELDKLPSSLVNVFVGEQATFALSGGDEYELCFTVPDS
NKAKLEALAYIGVDTYCVGOVRALSPQNKKRIFQDGNPIDITIQSGDFHFK"
complement(5448..5882)
/gene="nusb"
/note="PM0730"
complement(5448..5882)
/gene="nusb"
/codon_start=1
/transl_table=11
/product="Nusb"
/protein_id="AAK02814.1"
/db_xref="GI:12721025"
/translation="MTEQKHEKKISPRRRARECAVQALYSWYVSQNSPAEIELNFMAE
QDLKGVDTAYFRRLLPROTAEVNDVADVNIMIPYLDREVSELDPIEKAILRLAVYELKPE
LDVPYKVINEAIEVAKVFGAEDSHKYVNGVLDKVA PVLRSK"
complement(5888..6361)
/gene="ribh"
/note="PM0731"
complement(5888..6361)
/gene="ribh"
/codon_start=1
/transl_table=11
/product="Ribh"
/protein_id="AAK02815.1"
/db_xref="GI:12721026"
/translation="MKVLEGAAVAPNAKIYVVIARFNSFINESLLEGALDAKRLQGV
KEENTITVLPVPGAVELPLVARRLAESKKYDGIYALGTIRGTAHFHYVAGESSGLG
QVANNADIPVAFGVLTTEIEQAIERAGTKAGNKAGAAALVALEMVNLLAQIDAA"
6575..7786
/gene="tyrp_1"
/note="PM0732"
6575..7786
/gene="tyrp_1"
/codon_start=1
/transl_table=11
/product="tyrp"
/protein_id="AAK02816.1"
/db_xref="GI:12721027"
/translation="MKNKTFGSALLVAGTTIGAGMLAMPLTSAEMGFTYVTLILFILM

GLLSYALLFVEVQKAEKNAGIATLAQYFGLPGRILATLSLVWFMYALLSYVTV
 GSSLVGLPFLGEHATPISTVSFTVLGVFVIVSTGAVDMTLFELFMKLVAFVFL
 LMMFLPVNGNLAMPKDLRLTISASVFTSGFHFVITPSINSYLDGNTRRRLRIAL
 TETAPLVAYIIWOLATHVGFPOQAFVQLINTDPTLNGLTITATYQATESALISAMRL
 FFTALITSTFLGSLVFDCLYDLKLRVKIKTRVSYLSGLTLFLPLPLFALFYPEGFWM
 ALGYAGQMTFVGLVLPVGMARARKRYPDLPYRVGGNLLFGALLILGLIMNVFL
 IRAGVLPVAVIG.
 complement(7842, .8504)
 /gene="PM0733"
 complement(7842, .8504)
 /gene="PM0733"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAK02817.1"
 /db_xref="GI:12721028"
 /translation="MSQHTQQVLSAFHFSRAIRRYDPNKKVSEEDFHYILEGLRSL
 SVSGEPNHFILQRLQRLKAVSWGQTLDDASVILFTLAKKHARYDSVPYFHDIL
 LAKRGITTEVMQKALAIYERFQRDITQVLNDRNTLFDWACKQTYTALANMGTGAAMIG
 IDSCPIEGFDYCAVGNQLLSQTGADFPDNEYGVSVAVTFGYRAKEVKKPKSRPLTEITW
 LK".

Search completed: May 6, 2001, 16:29:35
Job time: 4796 sec

Db	3560	GTGATTTTACTTTGCTAGTGATGTCTTTGGTTTAAATTTAACCCCAACATGCGAG	3501
QY	445	cgcctggggcacttttgcctgatacttcgaaacgcctggcctggtttaccgcctcgttgc	504
Db	3500	ATCTGGAGCGCACTTTTGTGATCATTTATGGAAACCTTCTCTATTTTATGCCATTTC	3441
QY	505	agccctgtttgcctgcgcgaaatgcgcctggttatcaacgctctgcgaagtgcgattgat	564
Db	3440	ATTGTTTTTTCACGTCACAGCGCAAGCAGTTTTATGATCAATATAGCCGATATATTGAT	3381
QY	565	ggttttgcggggcgcttatttgcggatttggcattcatttgat	608
Db	3380	CATCTTTTCAGCGCGTATTTTCTTCGCTGTTGGTGCTATTTAAAT	3337

BASE COUNT	3070 a	2201 c	1937 g	2976 t
ORIGIN				

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 15:43:39 ; Search time 132.88 Seconds

(without alignments)
2715.058 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804

Perfect score: 618

Sequence: 1 atdttgatgtattttctcac.....ttcatttgattatttcgcgg 618

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0401.*

- 1: /SID52/gcgdata/geneseq/geneseqn/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseqn/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseqn/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseqn/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseqn/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseqn/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseqn/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseqn/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseqn/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseqn/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseqn/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseqn/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseqn/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseqn/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseqn/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseqn/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseqn/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseqn/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseqn/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseqn/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseqn/NA2001.DAT.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	618	100.0	840	21	A48443
2	54.8	8.9	672	21	A52688
C 3	36.2	5.9	8387	20	X84330
4	34.8	5.6	639	21	A52689
C 5	34.8	5.6	10732	21	A10594
C 6	33.4	5.4	1668	21	A61501
7	33.4	5.4	1668	21	A61502
C 8	32.4	5.2	565	21	C75485
C 9	32	5.2	3288	18	T68841
10	31.6	5.1	2178	21	Z53939
C 11	31.6	5.1	5059	20	X84332
					E. coli L-threonin
					Escherichia coli y
					Stealth virus nucl
					Escherichia coli y
					Gene encoding a su
					A. vitis hypersens
					A. vitis hypersens
					Human OREF ORP1040
					Photorhabdus lumin
					Neisseria gonorrhoe
					Stealth virus nucl

12	31.2	5.0	903	11	Q06352
13	31	5.0	114955	20	X53491
14	30.8	5.0	1535	22	A89040
15	30.8	5.0	16885	17	T33535
C 16	30.6	5.0	4296	21	Z54326
17	30.6	5.0	16526	21	A81472
C 18	30.6	5.0	172325	21	F21613
C 19	30.6	5.0	349980	21	F21612
C 20	30.6	5.0	837096	21	A81489
C 21	30.4	4.9	3288	19	V29932
C 22	30.4	4.9	92934	21	A81473
C 23	30.4	4.9	349980	21	F21544
C 24	30.4	4.9	1244	20	V72577
C 25	30.2	4.9	1244	20	V72578
C 26	30.2	4.9	3131	19	V30456
C 27	30.2	4.9	534720	19	V30458
C 28	30.2	4.9	536165	19	V30459
C 29	30	4.9	2163	21	Z53940
30	30	4.9	50925	21	A81487
31	30	4.9	349980	21	F21610
C 32	29.8	4.8	516	21	A79643
C 33	29.8	4.8	885	20	X80671
34	29.4	4.8	1065	21	A29323
35	29.4	4.8	1065	21	A29325
36	29.4	4.8	1195	21	A96225
37	29.2	4.7	1059	18	T73345
38	29.2	4.7	1100	21	A50040
39	29.2	4.7	2028	18	T73350
C 40	29	4.7	852	21	C43068
41	29	4.7	1934	17	T42302
42	29	4.7	38734	20	Z32020
43	29	4.7	38734	22	C90077
44	28.8	4.7	1116	21	A66028
45	28.8	4.7	4760	20	X02056

ALIGNMENTS

RESULT	1
A48443	ID A48443 standard; DNA; 840 BP.
XX	AC A48443;
XX	XX
DT	08-SEP-2000 (first entry)
DE	E. coli L-threonine resistance gene, rhtC.
XX	XX
KW	L-threonine resistance; L-threonine synthesis; rhtC;
KW	L-homoserine; L-valine; L-leucine; ds.
XX	XX
OS	Escherichia coli.
XX	XX
PH	Key Location/Qualifiers
FT	CDS 187..807
FT	/*tag= a
FT	/product= "RhtC"
XX	XX
PN	EP1013765-A1.
XX	XX
PD	28-JUN-2000.
XX	XX
PF	20-DEC-1999; 99EP-0125406.
XX	XX
PR	23-DEC-1998; 98RU-0123511.
XX	XX
PA	(AJIN) AJINOMOTO KK.
XX	XX
PI	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
XX	XX
DR	WPI: 2000-414602/36.
DR	P-PSDB: Y99598.

Modified carboxyl
Human adenosine A1
Mycobacterium tube
BCG deletion regio
Neisseria meningit
N. meningitidis pa
Neisseria meningit
Neisseria meningit
N. meningitidis pa
tcaA gene from the
N. meningitidis pa
Neisseria meningit
Mycobacterium tube
Mycobacterium tube
Human thiazide-sen
Rhizobium species
Rhizobium species
Neisseria meningit
N. meningitidis pa
Neisseria meningit
Pinus radiata cell
Clone if87_1 encod
Glycine max isofla
Glycine max isofla
CDNA encoding a ma
DNA encoding lipas
DNA encoding Chlam
Lipase gene expres
Arabidopsis thalia
Maize ribosomal in
Human METH1 relate
AL021529 cDNA clon
E. coli proliferat
Glycine max protop

XX Novel Escherichia bacterium having enhanced L-threonine resistance due
PT to enhanced RhtC protein activity, used to produce L-threonine,
PT L-homoserine, L-valine and L-leucine -
XX
XX Claim 9; Page 14-15; 24pp; English.
XX
XX The present sequence is the L-threonine resistance gene, rhtC, from
CC Escherichia coli. This sequence may be used to impart L-threonine
CC resistance on E. coli bacteria, which would be useful for producing
CC a high yield of L-threonine. L-threonine resistance means that the
CC bacteria will be able to grow on a minimal medium containing
CC L-threonine at a concentration at which the corresponding wild-type
CC strain would not grow. Since the transformed bacteria can grow on the
CC minimal medium, it can synthesise L-threonine, which accumulates. The
CC accumulated amino acids can then be removed from the culture medium.
CC The bacterium of the present invention may also be used to synthesise
CC L-homoserine, L-valine and L-leucine at increased levels.
XX
XX Sequence 840 BP; 165 A; 192 C; 248 G; 235 T; 0 other;

Query Match 100.0%; Score 618; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 4.1e-193;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 atgttgatgtattttctaccgtcgccatggtgcacattgtggcgcttatgagcccggt 60
Db 187 atgttgatgtattttctaccgtcgccatggtgcacattgtggcgcttatgagcccggt 246
QY 61 cccgattctttttctctcagaccgtgtcagtcgtccctcgttaaaagacgatgatg 120
Db 247 cccgattctttttctctcagaccgtgtcagtcgtccctcgttaaaagacgatgatg 306
QY 121 ggcgtctgggcaattaccctgcggcgtaattgttggcggtggatggcgcttggcgctg 180
Db 307 ggcgtctgggcaattaccctgcggcgtaattgttggcggtggatggcgcttggcgctg 366
QY 181 catttgattatcgaaaataagcctggctgcatacagctgattatgtggcggtgagctg 240
Db 367 catttgattatcgaaaataagcctggctgcatacagctgattatgtggcggtgagctg 426
QY 241 tatctctgtgattgggttaccagatgctacgtgtgtgcaactgaaaagagcggtttct 300
Db 427 tatctctgtgattgggttaccagatgctacgtgtgtgcaactgaaaagagcggtttct 486
QY 301 gcaactgcgccacagctcagctgcgaaagtggcgcaagtttctgaaaggttactg 360
Db 487 gcaactgcgccacagctcagctgcgaaagtggcgcaagtttctgaaaggttactg 546
QY 361 accaatctcgtaataccgaaagcattatctactttggctggctggtctcattgtttgtc 420
Db 547 accaatctcgtaataccgaaagcattatctactttggctggctggtctcattgtttgtc 606
QY 421 ggtgataacgttggcaactaccgcgcgtgggcatttttggcgctgattatgcgaaacg 480
Db 607 ggtgataacgttggcaactaccgcgcgtgggcatttttggcgctgattatgcgaaacg 666
QY 481 ctggcggtgtttaccgctgttccagcctgtttgcccctgcgcaaatgcgcgtggttat 540
Db 667 ctggcggtgtttaccgctgttccagcctgtttgcccctgcgcaaatgcgcgtggttat 726
QY 541 caacgtctggcgaagtggatgatgttttgcggggcggtatttgcgggatttggcatt 600
Db 727 caacgtctggcgaagtggatgatgttttgcggggcggtatttgcgggatttggcatt 786
QY 601 catttgattatttcgcgg 618
Db 787 catttgattatttcgcgg 804

RESULT 2
A52688

ID XX A52688 standard; DNA; 672 BP.
AC A52688;
DT 03-JAN-2001 (first entry)
XX DE Escherichia coli yahn gene.
XX KW E. coli; yahn gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
FT CDS 1..672
FT /*tag= a
FT /product= "yahn"
XX EPI016710-A2.
XX PD 05-JUL-2000.
XX PF 17-DEC-1999; 99EP-0125263.
XX PR 30-DEC-1998; 98RU-0124016.
XX DR 09-MAR-1999; 99RU-0104431.
XX PA (AJIN) AJINOMOTO CO INC.
XX PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV,
PI Tokhmakova IL;
XX WPI: 2000-414802/36.
XX P-PSDB; B01786.
XX Increased production of L-amino acids by an Escherichia bacterium
XX comprises increasing the expression amount of an L-amino acid excretion
XX protein -
XX Disclosure; Page 17-18; 29pp; English.
XX The present sequence is the yahn gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in lysine, glutamic acid and proline is
CC achieved if multiple copies of the gene are transfected into a bacterium.
CC The bacterium used is E. coli.
XX
XX Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;
Query Match 8.9%; Score 54.8; DB 21; Length 672;
Best Local Similarity 45.8%; Pred. No. 4.5e-08;
Matches 265; Conservative 0; Mismatches 307; Indels 6; Gaps 2;
QY 32 tgcacattgtggcgcttatgagcccggtcccgattctttttgtctctcagacgctg 91
Db 77 tgttcggtattacttttttaataccgggagccaatctcttggtagtaacaaccagcc 136
QY 92 tcaatcgttcccgtaaaagacgatgatggcgctgctggcattaccctgcggtaatgg 151
Db 137 tgggtccggtcgacgcgcgaggggtgctgacccggcggtggcgctggcgatgcat 196
QY 152 ttggggtgggattgcgctgcttggcgctgattgattatcgaaaaaatggcgctgctgc 211
Db 197 ttatttcgggttgggtttgttggcttctcaacgctaattacgcagtgtagagagattt 256
QY 212 atacgctgattatgttggcggtggcgctgctatctctgctgggttaccagatgctac 271
Db 257 ttctgcttatcaagaatcgtcgcggtcttatctctt---atggtttcgttggcgagc 312
QY 272 gtggtgcactgaaaaaagagcggttttctcaactgcccacaggtcgcagctggcgaaaa 331

Db 313 atgcccgcagtcacacaaatgagacacactaacacacgattagc--gccccct 370
 QY 332 gtggcgcagtttctgaaagttaactgacaaatctcgttaactcgaagcagattatct 391
 Db 371 ggtatgtcttttttcgcgcggtattactacccgatctcttaaccgcgaacacggtttat 430
 QY 392 actttggcgcggtgtctctatgttttcggtgataacggttgacactacgcgcgctggg 451
 Db 431 tttttcagatttttcacgataacataatgcgaaacacacacacggttttaa 490
 QY 452 gcaattttgcgctgatcattgtcgaacgcgtggcgtgtgttaaccgtgtgcagcctgt 511
 Db 491 tggcctggcgcggtgattgtcgtgcatacaattatctggcgagttttcttagcagcgct 550
 QY 512 ttgcctgcgcgaatgtcgcgtggtttatcaacgtctgcgaagtggtgattgaggtttg 571
 Db 551 ttctttgcccgcgtgtgcgtgctgtctatggcgtatgcgaacgcgttgccagtcgggtta 610
 QY 572 ccgggcgcttattgcgcggtttggcattcattgatt 609
 Db 611 ttggtgcaattggtgtattcgcgtacgcctgatt 648

RESULT 3

X84330/c
 ID X84330 standard; DNA; 8387 BP.
 AC X84330;
 DT 08-SEP-1999 (first entry)
 DE Stealth virus nucleic acid clone, SEQ ID NO: 22.
 XX Stealth virus; detection; diagnosis; infection; ss.
 KW Stealth virus.
 OS
 FH Key Location/Qualifiers
 FT misc_difference 8118 /*tag= a
 FT /*note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT misc_feature 8130..8133 /*tag= b
 FT /*note= "these nucleotides are represented as * in the
 FT specification, and are included to maintain the
 FT base numbering given in the specification"
 FT misc_difference 8157 /*tag= c
 FT /*note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT misc_difference 8275 /*tag= d
 FT /*note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 PN W09934019-A1.
 XX
 PD 08-JUL-1999.
 XX
 PF 30-DEC-1998; 98WO-US27744.
 XX
 PR 30-DEC-1997; 97US-0001184.
 XX
 PA (MART/) MARTIN W J.
 PI Martin WJ;
 XX WPI; 1999-405521/34.
 DR

XX Novel strains of stealth virus
 PF Claim 19; Page 61-64; 95pp; English.
 XX
 CC This sequence represents a stealth virus nucleic acid clone. The
 CC invention relates to a method of detecting and characterising a stealth
 CC virus by reacting a sample suspected of containing a stealth virus with a
 CC probe from a known stealth virus and sequencing the resultant isolated
 CC nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
 CC from a sample suspected of containing a stealth virus, e.g. a culture of
 CC cells showing a viral cytopathic effect; (b) testing the reactivity of
 CC the isolated DNA or RNA with a molecular probe that contains at least 18
 CC or more contiguous nucleotides identical to sequence previously
 CC identified from a stealth virus; and, optionally (c) sequencing the
 CC isolated DNA or RNA molecules that react with the probe. The method is
 CC used to detect stealth virus in a biological product, food or in the
 CC environment. The method is also used to evaluate agents for their
 CC inhibitory or stimulatory effects on stealth virus replication and to
 CC determine capacity of the virus to recombine with and potentially alter
 CC the nucleic acid sequences of a cell or bacterium.
 XX
 SQ Sequence 8387 BP; 1807 A; 2320 C; 2218 G; 1844 T; 198 other;
 Query Match 5.9%; Score 36.2; DB 20; Length 8387;
 Best Local Similarity 23.5%; Pred. No. 0.2;
 Matches 77; Conservative 87; Mismatches 163; Indels 0; Gaps 0;
 QY 150 ggtttggcgtgattgcgtgctgctgcctgcattgattatcgaaaaaatgcctgct 209
 Db 8346 SNKHTSYAAMVKGWATKYNRYTKKSANKGNATKYNRYTKKSANKGHYVWTASTRYA 8287
 QY 210 gcatacgcgtgattatggtggcgtggtgcctgtatctctgctggtggttaccagatgt 269
 Db 8286 NAYNKMHDGVTNTTHVGWNTCHAAAGVHVGGDHVWASYGKNAYTKNDWTKNYATKND 8227
 QY 270 acgtgggtgcactgaaagagcggttctgcactgcacagtcgacagtcgagtcgcaaa 329
 Db 8226 WVGVAHDNATKNDWATKNDWATKNDWATKNDWATKNDWATKNDWATKNDWATKND 8167
 QY 330 aagtggcgcgcttctcgtgaaagtttactgacaaatctcgttaattccgaaagcattat 389
 Db 8166 HNAYTKNDWANDTYCTYTATHTDKTYNTVGHNNNNNDKNTYSABWRTDWGAATTCGT 8107
 QY 390 ctacttggcgtggttctcattgtttgctggtgataacggttgacactacgcgcgtg 449
 Db 8106 CGATCTGCTCGAGAGTCCGCGTGTGATTTCGTCGTTGATGTGAGAAATAATGCCGCGTTC 8047
 QY 450 gggcatttttgcgctgatcatgttcga 476
 Db 8046 GCGCACTAATCCGAGTTCACAGSCGA 8020
 RESULT 4
 A52689
 ID A52689 standard; DNA; 639 BP.
 XX
 AC A52689;
 DT 03-JAN-2001 (first entry)
 XX
 DE Escherichia coli yeas gene.
 XX
 KW E. coli; yeas gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein; ds.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..639 /*tag= a
 FT /*product= "yeas"
 FT

SQ Sequence 3288 BP; 894 A; 793 C; 768 G; 833 T; 0 other;

Query Match 5.2%; Score 32; DB 18; Length 3288;
Best Local Similarity 46.4%; Pred. No. 3;
Matches 104; Conservative 0; Mismatches 120; Indels

Qy	159	tgggattgagctgttgctgcctcatttgaattatcgaaaaatggctggtgcatacgt	218
Db	2728	TGGCATTTGCTATGGCTAAACAGGGTTTGATTATCATGCCAGTAGTTGCCAATGTAAAAG	2669
Qy	219	gattatgggtggcgggtggcctgatctctgctgcatgggtaccagatgctaagtggtagc	278
Db	2568	GATTATACACTACCAGATGTGTGATTTTCTCCAGCAGAAGGTGCAGATACCCATCACATTGC	2609
Qy	279	actgaaaaaaggcgggtttctgcacctgcgccacaggctcagctggcgaaaasgtggcg	338
Db	2508	TCITTANTGCCGAACCGGGCGCTTACGGCCGCGAGCACTTCTGCTGAACGCCGAATCCA	2549
Qy	339	cagttctcgaaggtttactgaccaatctcgtcaatccgaag	382
Db	2548	TTGTTTTCTTTAGTGGCAGCTTCTACATTATTCAGCTGTGCACAAACG	2505

RESULT 10

233939
ID 253939 standard; DNA; 2178 bp.

XX 253939: AC

XX
DT 21-MAR-2000 (first entry)

XX
DE
Notarische Beglaubigung

XX
XX

KW antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;

XX
антисцелестрат, генетерау, аз.
XX

US
Neisseria gonorrhoeae.
XX

PN
W09957280-A2.
yy

PD 11-NOV-1999.

PF 30-APR-1999; 99WO-US09346.

PR 01-MAY-1998; 98US-0083758.

PR 31-JUL-1998; 980S-0094869.
PR 02-SEP-1998. 987S-0098994

PR 02-SEP-1998; 98US-0099062.
 00-OCT-1998; 98US-0100740
 00-DEC-1998; 98US-0102740

PR 09-OCT-1998; 98US-0103794.

PR 25-FEB-1999; 99US-0121528.

XX PA (CHIR) CHIRON CORP.

PA (GENO-) INST GENOMIC RES.
XX

PI Fraser C, Galeotti C, Grandi G, Hickey E, Massignani V, Mora M;

PI Tettelin H, Venter JC;

DR WPI; 2000-062150/05.

OK F-PSUB; 1/31/77.
XX

PT Novel Neisserial polypeptides predicted to be useful antigens for vaccines and diagnostics -

XX
DS
C]aim 7. Dno 010-010. 145300. 145300.

XX

CC novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides and

CO POPULATIONS: 23433/ CO 23437/0 and 234310 CO 23347/3 represent FOR

primers used in the exemplification of the present invention. The polypeptides, the polynucleotides, antibodies and compositions of the invention can be used as vaccines, as diagnostic reagents, and as immunogenic compositions. The polypeptides can be used in the manufacture of medicaments for treating or preventing infection due to *Neisseria* bacteria (e.g. meningitis and septicaemia), to detect the presence of *Neisseria* bacteria, or to raise antibodies. They may also be used to screen for agonists or antagonists, which may themselves have use as antibacterial agents. The polynucleotides of the invention may also be used in gene therapy protocols.

Sequence 2178 BP; 521 A; 589 C; 622 G; 446 T; 0 other;

Query Match 5.1%; Score 31.6; DB 21; Length 2178;
Best Local Similarity 58.5%; Pred. No. 3.4;
Matches 55; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 40 gtggcgcttagagcccggtcccgattctctttttgtctctctagacccgctgcagtcgt 99
Db 1099 gtcgactgatcaccccttgcgcttttgggtattgctgccgtgcgctgcgtcgtg 1158
Qy 100 tcccgtaaagaagcgtatgatggcgctgcgggca 133
Db 1159 qcacccctcccgctattgatgcagatgcgca 1192

RESULT 11
X84332/c
ID X84332 standard; DNA; 5059 BP.

DT	08-SEP-1999	(first entry)
XX		
DE	Stealth virus nucleic acid clone,	SEQ ID NO: 24.
XX		
KW	Stealth virus; detection; diagnosis;	infection; ss
XX		
OS	Stealth virus.	

Key	Location/Qualifiers
misc_difference 3605	a "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3610	b "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3615	c "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3630	d "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3631	e "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3638	f "this nucleotide is represented as a * in the specification, and is included to maintain the base numbering given in the specification"
misc_difference 3641	

[illegible][illegible]


```
XX 17-SEP-1998; 98WO-US19419.
XX
XX 09-JUN-1998; 98US-0093972.
XX
XX 17-SEP-1997; 97US-0059160.
XX
XX (UYEC-) UNIV EAST CAROLINA.
XX
XX Nyce JW;
XX
XX WPI; 1999-229400/19.
XX
XX New antisense oligonucleotides used in treatment of, e.g. pulmonary
XX vasoconstriction
XX
XX Disclosure: Page 37; 120pp; English.
XX
XX The specification describes antisense oligonucleotides (X52869-X55271)
XX directed against at least 2 mRNAs selected from target genes, coding and
XX non-coding regions of RNAs corresponding to target genes, gene
XX initiation codons, genomic flanking regions, intron-exon borders, the
XX 5'-end, the 3'-end and the juxta-section between coding and non-coding
XX regions and all segments of RNAs encoding proteins associated with one
XX or more diseases, conditions or mixtures. The antisense oligonucleotides
XX may be derived from sequences X5272-74. These multiple target
XX oligonucleotides (specifically X55180-271) can be used for the antisense
XX treatment of diseases and conditions. Typical diseases and conditions
XX are those associated with impaired respiration and inflammation,
XX including lung diseases, pulmonary vasoconstriction, inflammation,
XX allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
XX respiratory distress syndrome, pain, cystic fibrosis, pulmonary
XX hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
XX pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
XX carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
XX cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
XX metastases, as well as all types of cancers which may metastasize or have
XX metastasized to the lungs, including breast and prostate cancer.
XX
XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;
XX
XX Query Match 5.0%; Score 31; DB 20; Length 114955;
XX Best Local Similarity 57.3%; Pred. No. 37;
XX Matches 47; Conservative 2; Mismatches 33; Indels 0; Gaps 0;
XX
XX 6 gatgtatttcacgcgtgcacattgtggcgttatgtgaccccggtcccg 65
XX
XX 74086 ggtgtttctcttggttgctgtcgtcgtgctcnnndngctccgctcccg 74145
XX
XX 66 ttctctttttgtctctcagacc 87
XX
XX 74146 gtctcgtcgtctgtgcctcc 74167
XX
XX RESULT 14
XX A89040
XX ID A89040 standard; DNA; 1536 BP.
XX
XX AC A89040;
XX
XX
XX 05-MAR-2001 (first entry)
XX
XX Mycobacterium tuberculosis open reading frame mtb6, partial.
XX
XX MTB6; tuberculosis; BCG; vaccine; infection; diagnosis; ds.
XX
XX Mycobacterium tuberculosis.
XX
XX Key Location/Qualifiers
XX CDS 1..1536
XX FT /*tag= a
XX FT /partial
XX PN W09625519-A1.
XX
```

```
PN W0200066157-A1.
XX
XX 09-NOV-2000.
XX
XX 04-MAY-2000; 2000WO-US12257.
XX
XX 04-MAY-1999; 99US-0132505.
XX
XX (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
XX
XX Gennaro ML;
XX
XX WPI; 2001-007153/01.
XX
XX P-PSDB; B19847.
XX
XX Novel polypeptide encoded by open reading frames present in
XX Mycobacterium tuberculosis genome and not by the BCG strain of M.
XX bovis, useful as vaccine and for diagnosing tuberculosis infection -
XX
XX Claim 1; Fig 2; 35pp; English.
XX
XX The present sequence is that of open reading frame (ORF) mtb6
XX (partial) of Mycobacterium tuberculosis. This is 1 of 8 ORFs (see
XX A89035-42) identified as being present in the genome of M.
XX tuberculosis but absent from the genome of the BCG strain of
XX Mycobacterium bovis. The proteins encoded by these ORFs, i.e.
XX MTB6L-8 (see B19842-49) represent reagents that are useful in
XX for discriminating between M. tuberculosis and BCG and, in particular,
XX for diagnostic methods which discriminate between exposure of a
XX subject to M. tuberculosis and vaccination with BCG. The invention
XX features these MTB6 polypeptides, functional fragments of them, DNA
XX encoding them, vectors, transformed cells, and diagnostic,
XX therapeutic, and prophylactic (vaccine) methods, including genetic
XX vaccination methods.
XX
XX Sequence 1536 BP; 224 A; 473 C; 519 G; 320 T; 0 other;
XX
XX Query Match 5.0%; Score 30.8; DB 22; Length 1536;
XX Best Local Similarity 58.9%; Pred. No. 5.2;
XX Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;
XX
XX 440 ccgcgcgcgtggggcatttttcgctgatcattgcgaacgcgtggcgtgttacgcgtcg 499
XX
XX 854 ccgcgcgcgtggggatcgatcgcgtggcgtgttcattgtgacgaatgcgccaagctgaccgtcg 913
XX
XX 500 ttgccagcctgttgcctgcgcgcaaatgc 529
XX
XX 914 ccgtgcgcgcgtgcgtgcgcgcgattc 943
XX
XX RESULT 15
XX T33535
XX ID T33535 standard; DNA; 16885 BP.
XX
XX AC T33535;
XX
XX
XX 15-FEB-1998 (first entry)
XX
XX BCG deletion region 1 and flanking sequences.
XX
XX BCG delta 1; virulence; avirulence; attenuation; gene deletion;
XX mycobacteria; vaccine; infection; marker; ss.
XX
XX Mycobacterium bovis strain BCG.
XX
XX Key Location/Qualifiers
XX misc_feature 2327..11126
XX FT /*tag= a
XX FT /note= "BCG delta 1 deletion region"
XX
XX PN W09625519-A1.
XX
```

```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; FLOOR
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-390-878-16

Query Match 5.0%; Score 30.8; DB 1; Length 16885;
Best Local Similarity 58.9%; Pred. No. 3.8;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 440 ccgcgcgtcggggcatttttgcgtgcattgtcgaaacgtgcgtgtttaccgtcg 499
DB 7930 CCGCGGGGGGATCGCATCGGGTGTTCATGTGACGAATCGCGCCAACTGACCGTCG 7989
QY 500 ttgcagcgtgtttgcctcgcgaaatgc 529
DB 7990 CGGTGCGCGGATCGCGTGCCTGCGCGGATTC 8019

RESULT 3
US-09-067-626-1/G
; Sequence 1, Application US/09067626
; Patent No. 6177086
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; APPLICANT: Nathan, Carl F.
; APPLICANT: Ehrt, Sabine
; TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
; TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,626
```

```
;
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 60/045,688
; APPLICATION NUMBER: 06-MAY-1997
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-067-626-1

Query Match 4.9%; Score 30.2; DB 4; Length 1244;
Best Local Similarity 49.7%; Pred. No. 1.4;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 185 tgattatcgaaaaaatggcctgctgcacgtgattatgttggtggcgtggtgtatc 244
DB 595 TGCTCAGCGACGAGATCCCGTGTGCGATCCGCGGTGCGCGGCTGCGACCCGTGCATC 536
QY 245 tctgtggtggtgttaccagatcgtcgtgtgtcactgaaaaaaggcggtttctgcac 304
DB 535 TCGGCGCACCCGTGACCATGCGCGCGCCGAGGACGAGCTCGCGCGGGGACGCCACG 476
QY 305 ctgcgcacagtcgagtcgtgcgcaaaagtggcgcc 339
DB 475 CCGACTGCGCGATGCTGCTGCGCGCGTGTCCGCAC 441

RESULT 4
US-09-067-626-3
; Sequence 3, Application US/09067626
; Patent No. 6177086
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; APPLICANT: Nathan, Carl F.
; APPLICANT: Ehrt, Sabine
; TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
; TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,626
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/045,688
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
```

```
; REFERENCE/DOCKET NUMBER: 19603/491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-067-626-3

Query Match          4.9%; Score 30.2; DB 4; Length 1244;
Best Local Similarity 49.7%; Pred. No. 1.4;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 185 tgattatgaaaaatggcctggctgcatacgtgattatgtgtggcggtggcctgtatc 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 650 TGTCTAGCGACGAGATCCCGTGGTCCGATCCGCGCTGCGGGCGGCTGCGAGCCTGCATC 709
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 tctgtgattgggttaccagatgtctacgtgtgcaactgaaaaaaggcggtttctgcac 304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 710 TCGCGGACCCGTCACGATGGCCGCCGCGGAGGCGACGTCGCGCGGAGGCGGACG 769
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 305 ctgcgccacagatcgagctggcgaaaaadtgggcgc 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 770 CCGACTGGCGGATGGTGGCGCGCGGTGTCGCGCAC 804
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-07-941-651-1
; Sequence 1, Application US/07941651
; Patent No. 5332808
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; ADDRESSEE: Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5332808th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,651
; FILING DATE: 19920908
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 990..1826
US-07-941-651-1

Query Match          4.7%; Score 29; DB 1; Length 1934;
Best Local Similarity 48.0%; Pred. No. 4.6;
Matches 83; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 218 tgattatggggcggtggcctgtatctctgtctgggatgtttaccagatgtctacgtgtg 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1576 TGATCATGGTATGCGAGGGGCTGCGGTCTTCTACCGCTGTCCCGCAAGGTAGACGAGGGGT 1635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 278 cactgaaaaaaggcggtttctgcaacctgcaccacagctcgagctgcgaaaaagtgggc 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1636 TCAAGAAAGCCGCAAGCGGTGACCATATCGCGCTGTGAGGGGGAAGCAGGTGCAGAAATGGG 1695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 gcagtttctgaaagggtttactgaccaatctcgctaatactccgaaagcgattatc 390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1696 ACAGGATCTCGAAAGCGGTCTTCAGGTGGGCCGCTGACCCGACCGCTGAGATC 1748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-08-279-996-1
; Sequence 1, Application US/08279996
; Patent No. 5552140
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; ADDRESSEE: Gibson
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5552140th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/279,996
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,651
; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 990..1826
```


Db 477 GGTATTCAAAATGAAATGAAGTTTATTAAGTATGCTTCAATTAACACGAGA 536

QY 266 tgctacgtggtgcactgaaaaa 287

Db 537 AACTTGTATTATTTCTAAAGAA 558

RESULT 9

US-09-092-409-53

; Sequence 53, Application US/09092409

; Patent No. 6159478

; GENERAL INFORMATION:

; APPLICANT: Haanes, Elizabeth J.

; APPLICANT: Frank, Rexann S.

; TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES

; NUMBER OF SEQUENCES: 92

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sheridan Ross & McIntosh

; STREET: 1700 Lincoln Street, Suite 3500

; CITY: Denver

; STATE: Colorado

; COUNTRY: U.S.A.

; ZIP: 80203

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/092,409

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/680,726

; FILING DATE: 12-JUL-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Connell, Gary J.

; REGISTRATION NUMBER: 32,020

; REFERENCE/DOCKET NUMBER: 2618-46-C1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (303) 863-9700

; TELEFAX: (303) 863-0223

; INFORMATION FOR SEQ ID NO: 53:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1038 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..1035

US-09-092-409-53

Query Match 4.6%; Score 28.4; DB 4; Length 1038;

Best Local Similarity 50.0%; Pred. No. 5.2;

Matches 71; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 146 taatggttggcggtgattgcgtgcttgccctgattgatacgaataatggcct 205

Db 417 TGATGAATGGGATTAGTTTACGTGCCCGCTCAATTAATCAAGTCAATATAGACG 476

QY 206 ggcgcatacgcgtgattggtggcggtggcctgtatctctgctggatgggtaccaga 265

Db 477 GGTATTCAAAATGAAATGAAGTTTATTAAGTATGCTTCAATTAACACGAGA 536

QY 266 tgctacgtggtgcactgaaaaa 287

Db 537 AACTTGTATTATTTCTAAAGAA 558

RESULT 10

US-08-220-151-18

; Sequence 18, Application US/08220151

; Patent No. 5529780

; GENERAL INFORMATION:

; APPLICANT: Paolletti, Enzo

; APPLICANT: Limbach, Keith J.

; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF

; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC AND gD AND USES THEREFOR

; NUMBER OF SEQUENCES: 91

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford

; STREET: 530 Fifth Avenue

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/220,151

; FILING DATE: 30-MAR-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Frommer, William S.

; REGISTRATION NUMBER: 25,506

; REFERENCE/DOCKET NUMBER: 454310-2540

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 840-3333

; TELEFAX: (212) 840-0712

; TELEX: 425066 CURTMS

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1320 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-220-151-18

Query Match 4.6%; Score 28.4; DB 1; Length 1320;

Best Local Similarity 50.0%; Pred. No. 5.9;

Matches 71; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 146 taatggttggcggtgattgcgtgcttgccctgattgatacgaataatggcct 205

Db 617 TGATGAATGGGATTAGTTTACGTGCCCGCTCAATTAATCAAGTCAATATAGACG 676

QY 206 ggcgcatacgcgtgattggtggcggtggcctgtatctctgctggatgggtaccaga 265

Db 677 GGTATTCAAAATGAAATGAAGTTTATTAAGTATGCTTCAATTAACACGAGA 736

QY 266 tgctacgtggtgcactgaaaaa 287

Db 737 AACTTGTATTATTTCTAAGAA 758

RESULT 11

US-08-413-118-18

; Sequence 18, Application US/08413118

; Patent No. 5688920

; GENERAL INFORMATION:

; APPLICANT: PAOLETTI, ENZO

; APPLICANT: LIMBACH, KEITH J.

; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF

; TITLE OF INVENTION: CANINE HERPESVIRUS gB, gC, AND gD AND USES THEREFOR

; NUMBER OF SEQUENCES: 128

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.

; STREET: 530 FIFTH AVENUE, 25TH FLOOR

Qy 266 tgctacgtggtgcactgaaaaa 287
|| | | | | | |
Db 5664 AACTGTGTATTTTCTAAGAA 5685

Search completed: May 6, 2001, 16:25:06
Job time: 2647 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 14:41:24 ; Search time 1151.46 Seconds
(without alignments)
4688.750 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804
Perfect score: 618
Sequence: 1 atgttgatgtattttctcac.....ttcatttgattatttcggg 618

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST:*
- 1: gb_est1.*
 - 2: gb_est2.*
 - 3: gb_est3.*
 - 4: gb_est4.*
 - 5: gb_est5.*
 - 6: gb_est6.*
 - 7: gb_est7.*
 - 8: gb_est8.*
 - 9: gb_est9.*
 - 10: gb_est10.*
 - 11: gb_est11.*
 - 12: gb_est12.*
 - 13: gb_est13.*
 - 14: gb_est14.*
 - 15: gb_est15.*
 - 16: gb_est16.*
 - 17: gb_est17.*
 - 18: gb_est18.*
 - 19: gb_est19.*
 - 20: gb_est20.*
 - 21: gb_est21.*
 - 22: gb_est22.*
 - 23: gb_est23.*
 - 24: gb_est24.*
 - 25: gb_est25.*
 - 26: gb_est26.*
 - 27: gb_est27.*
 - 28: gb_est28.*
 - 29: gb_est29.*
 - 30: gb_est30.*
 - 31: gb_est31.*
 - 32: gb_est32.*
 - 33: gb_est33.*
 - 34: gb_est34.*
 - 35: gb_est35.*
 - 36: gb_est36.*
 - 37: gb_est37.*
 - 38: gb_est38.*
 - 39: gb_est39.*
 - 40: gb_est40.*
 - 41: gb_est41.*
 - 42: gb_est42.*
 - 43: gb_est43.*
 - 44: gb_est44.*
 - 45: gb_est45.*
 - 46: gb_est46.*
 - 47: gb_est47.*
- em_esthum10.*
- 44: em_esthum10.*
 - 45: em_esthum11.*
 - 46: em_esthum12.*
 - 47: em_esthum13.*
 - 48: em_esthum14.*
 - 49: em_esthum15.*
 - 50: em_esthum16.*
 - 51: em_esthum17.*
 - 52: em_esthum18.*
 - 53: em_esthum19.*
 - 54: em_esthum20.*
 - 55: em_esthum21.*
 - 56: em_esthum22.*
 - 57: em_esthum23.*
 - 58: em_esthum24.*
 - 59: em_esthum25.*
 - 60: em_esthum26.*
 - 61: em_esthum27.*
 - 62: em_esthum28.*
 - 63: em_estin1.*
 - 64: em_estin2.*
 - 65: em_estin3.*
 - 66: em_estin4.*
 - 67: em_estin5.*
 - 68: em_estomi.*
 - 69: em_estom2.*
 - 70: em_estov1.*
 - 71: em_estov2.*
 - 72: em_estpl1.*
 - 73: em_estpl2.*
 - 74: em_estpl3.*
 - 75: em_estpl4.*
 - 76: em_estpl5.*
 - 77: em_estpl6.*
 - 78: em_estpl7.*
 - 79: em_estpl8.*
 - 80: em_estpl9.*
 - 81: em_estpl10.*
 - 82: em_estro1.*
 - 83: em_estro2.*
 - 84: em_estro3.*
 - 85: em_estro4.*
 - 86: em_estro5.*
 - 87: em_estro6.*
 - 88: em_estro7.*
 - 89: em_estro8.*
 - 90: em_estro9.*
 - 91: em_estro10.*
 - 92: em_estro11.*
 - 93: em_estro12.*
 - 94: em_estro13.*
 - 95: em_estro14.*
 - 96: em_estro15.*
 - 97: em_estro16.*
 - 98: em_estro17.*
 - 99: em_estro18.*
 - 100: em_estro19.*
 - 101: em_estro20.*
 - 102: gb_est25.*
 - 103: gb_est26.*
 - 104: gb_est27.*
 - 105: gb_est28.*
 - 106: gb_est29.*
 - 107: gb_est30.*
 - 108: gb_est31.*
 - 109: gb_est32.*
 - 110: gb_est33.*
 - 111: gb_est34.*
 - 112: gb_est35.*
 - 113: gb_est36.*
 - 114: gb_est37.*
 - 115: gb_est38.*
 - 116: gb_est39.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match		Query Length	DB ID	Description
		No.	Match			
1	38.6	6.2	922	229	CNS0073W	AL066784 Drosophil
C 2	38.4	6.2	425	204	AQ280480	AQ280480 C11B1-E1-
C 3	36.2	5.9	1108	139	BE793235	BE793235 601583674
4	35.8	5.8	977	229	CNS00JX7	AL076850 Drosophil
C 5	35.6	5.8	912	229	CNS006RB	AL065826 Drosophil
C 6	35.6	5.8	939	229	CNS00CNG	AL059400 Drosophil
C 7	35.2	5.7	914	231	CNS0300A	AL252739 Tetraodon
8	35.2	5.7	1016	231	CNS04744	AL277501 Tetraodon
9	35.2	5.7	1068	230	CNS028DO	AL185829 Tetraodon
10	35	5.7	1200	229	CNS01671	AL106392 Drosophil
11	34.6	5.6	835	150	BF621090	BF621090 HVSMea000
12	34.6	5.6	920	229	CNS0062R	AL061710 Drosophil
13	34.4	5.6	1101	229	CNS017GJ	AL108013 Drosophil
14	34.2	5.5	832	142	BE957648	BE957648 601653724
C 15	33.8	5.5	501	32	AV720117	AV720117 AV720117
C 16	33.6	5.4	268	119	AW721077	AW721077 833011D11
C 17	33.6	5.4	1355	164	BE196650	BE196650 HVSMea009
C 18	33.4	5.4	571	206	AQ438206	AQ438206 HS_5083_B


```

JOURNAL      Tetraodon nigroviridis DNA sequence
REFERENCE    Unpublished
AUTHORS      3 (bases 1 to 1068)
TITLE        Genoscope.
COMMENT      Direct Submission
             Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
             This sequence is a single read and was generated as part of a large
             scale clone-end sequencing project of the Tetraodon nigroviridis
             genome. For more information, please take a look at
             http://www.genoscope.cns.fr/tetraodon.
FEATURES     Location/Qualifiers
             source
               1..1068
               /organism="Tetraodon nigroviridis"
               /db_xref="taxon:99883"
               /clone="24L11"
               /clone_lib="G"
               /note="Genoscope sequence ID : COAG244CF06SP1-end :
               PUC-Orig"
BASE COUNT   251 a 301 c 272 g 237 t 7 others
ORIGIN
Query Match      5.7%; Score 35.2; DB 230; Length 1068;
Best Local Similarity 54.7%; Pred. No. 4.5;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 433 ggcaactacgcgcgtggtgggatttttgcgtgctgacattgtcgaacgctggcggtgtt 492
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 829 GCGTTCACCGCGCGCACGACGATTCCTCGCTCGACGCCCATACCCCTCGCGCTGCAG 888
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 493 accgtgttgccagcgtgttgcctgcccaatgcgcctggttatcaacgtctgcgcg 552
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 889 AGGGGTGTCCGCGTCCGCTACCCAGCGGTGGACGCCGCTTCACAGCGCCGCGCTAG 948
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 553 aagtgat 560
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 949 ACGTCGCT 956

RESULT 10
CNS01671
LOCUS      CNS01671      1200 bp      DNA      GSS      26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN15022 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION  AL106392
VERSION     AL106392.1 GI:5621686
KEYWORDS   GSS.
SOURCE     fruit fly.
ORGANISM   Plasmid Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 1200)
AUTHORS     Direct Submission
TITLE       Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBelOBAC11.
FEATURES     Location/Qualifiers
             source
               1..1200
               /organism="Drosophila melanogaster"
               /plasmid="pBelOBAC11"
               /db_xref="taxon:7227"
               /clone_lib="DrosBAC"
Tetraodon nigroviridis DNA sequence
Unpublished
3 (bases 1 to 1068)
Genoscope.
Direct Submission
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.
Location/Qualifiers
1..1068
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="24L11"
/clone_lib="G"
/note="Genoscope sequence ID : COAG244CF06SP1-end :
PUC-Orig"
BASE COUNT 251 a 301 c 272 g 237 t 7 others
ORIGIN
Query Match      5.7%; Score 35; DB 239; Length 1200;
Best Local Similarity 32.3%; Pred. No. 5.4;
Matches 70; Conservative 48; Mismatches 99; Indels 0; Gaps 0;

QY 397 ggctcggtgttcattgttgcgttgataaacgttggcaactaccgcgcgtgggcatt 456
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 698 GGGSCGCGCTGTCGSGGCTKSGCGSCCCYTTGTYTTRCTKCCGCGGKKTSTYTGTS 757
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 457 ttgcgctgcatcattgtcgaacgcgtgctgggtgtttaccgtcttccagctgtttgc 516
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 758 GTCCCGSGSTKBBTYYTKGSTBYCYTGGCCGCSBYTSCGCGCTBCGCCSTGKSCYCS 817
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 517 ctgcgcgaatgcgcgtggttatcaactctgcgaagtggattgatggtttgcgcgg 576
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 818 SYISGCGCGCTYCSGCCGSCCTYCSGCCGCTYCYGCGKGBKSCGCGCSYTCCTGKSBGG 877
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 577 gcgttattgcgcgatttgcgttcattcatttattt 613
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 878 GSGBGSTTKGBKBBBKYTYTGTYYTKCCTTT 914
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
BF621090
LOCUS      BF621090      835 bp      mRNA      EST      18-DEC-2000
DEFINITION HVSMEa0003P06f Hordeum vulgare seedling shoot EST library
            HVCdNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0003P06f,
            mRNA sequence.
ACCESSION  BF621090
VERSION     BF621090.1 GI:11884824
KEYWORDS   EST.
SOURCE     barley.
ORGANISM   Hordeum vulgare
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae;
            Hordeum.
REFERENCE   1 (bases 1 to 835)
AUTHORS     Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
            , Y., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
            , T., Saski, C., Schwartzbeck, J., Simmons, J., Chol, D.W., Main, D. and
            Wood, T.
            Development of a genetically and physically anchored EST resource
            for barley genomics
            Unpublished (2000)
            Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCTCACTAAAGG
            High quality sequence start: 8
            High quality sequence stop: 515.
            Location/Qualifiers
             source
               1..835
               /organism="Hordeum vulgare"
               /cultivar="Morex"
               /db_xref="taxon:4513"
               /clone="HVSMEa0003P06f"
               /clone_lib="Hordeum vulgare seedling shoot EST library
               HVCdNA0001 (Cold stress)"
               /tissue_type="Seedling shoot"
               /lab_host="TJC121"
               /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 232 a 229 c 158 g 216 t
ORIGIN

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:38:48 ; Search time 34.23 Seconds
(without alignments)
344.015 Million cell updates/sec

Title: US-09-466-935-4
Perfect score: 1054
Sequence: 1 MLMFLITVAMVHVALMSPG.....IDGFAGALFAGGIHLIISR 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401.*
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054	100.0	206	Y99598	E. coli L-threonin
2	261.5	24.8	223	B01786	Escherichia coli Y
3	190	18.0	212	B01787	Escherichia coli Y
4	179.5	17.0	229	B11637	A. vitis hyperseus
5	122	11.6	205	Y99597	E. coli L-homoseri
6	122	11.6	205	Y79298	E. coli RhtB prote
7	105	10.0	236	W37715	C. glutamicum Lys
8	103	9.8	211	B01789	Escherichia coli Y
9	101	9.6	195	B01788	Escherichia coli Y
10	94.5	9.0	130	B40407	Human ORFX ORF171
11	91	8.6	153	R87527	Mel-linked mlgA ge

12	91	8.6	153	20	W73358	S. colwelliana Mlg
13	86.5	8.2	431	17	R98004	Oestrogen-regulate
14	86.5	8.2	431	19	W34528	Protein encoded by
15	86	8.2	390	11	R05475	Chloramphenicol re
16	82.5	7.8	356	20	Y38893	Neisseria gonorrhoe
17	82	7.8	713	21	Y58582	Sorangium cellulos
18	81.5	7.7	439	21	Y39079	Arabidopsis thalia
19	81.5	7.7	444	21	G21141	Arabidopsis thalia
20	81.5	7.7	444	21	G39078	Arabidopsis thalia
21	81.5	7.7	555	21	G39077	Arabidopsis thalia
22	81.5	7.7	559	21	G21140	Arabidopsis thalia
23	81.5	7.7	579	21	G21139	Arabidopsis thalia
24	81.5	7.7	1170	20	W88447	Yeast NPC1 protein
25	81	7.7	381	11	R03463	GPL transglutamina
26	80.5	7.6	299	20	W90237	E. coli antibiotic
27	80.5	7.6	306	20	W90236	E. coli antibiotic
28	80	7.6	356	20	Y38890	Neisseria meningit
29	80	7.6	356	20	Y38891	Neisseria meningit
30	80	7.6	609	21	B08549	Amino acid sequenc
31	79.5	7.5	1307	17	R99255	Aspergillus flavus
32	78	7.4	1518	20	Y16099	Acetobacter xylinu
33	77	7.3	332	20	Y34112	Human GABA recepto
34	77	7.3	332	20	Y30312	Human GABA recepto
35	77	7.3	332	20	Y29662	Amino acid sequenc
36	77	7.3	859	20	Y34111	Human 7TM receptor
37	77	7.3	859	20	Y30311	Human GABA recepto
38	77	7.3	874	20	Y29661	Amino acid sequenc
39	77	7.3	898	20	Y14082	Human 7TM receptor
40	77	7.3	898	21	Y70326	Human GABABR2 prot
41	77	7.3	914	21	Y44344	Human gamma amino
42	77	7.3	941	20	Y29796	Protein-1 related
43	77	7.3	941	21	W90938	Human gamma-amino-
44	77	7.3	941	21	Y51928	Human GABA-B recep
45	77	7.3	941	21	Y70328	Human GABA-B-R2 re

ALIGNMENTS

RESULT	1
Y99598	
ID	Y99598 standard; Protein; 206 AA.
XX	
AC	Y99598;
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	E. coli L-threonine resistance protein, RhtC.
XX	
KW	L-threonine resistance; L-threonine synthesis; rhtC;
KW	L-homoserine; L-valine; L-leucine.
XX	
OS	Escherichia coli.
XX	
PN	EPI013765-Al.
XX	
PD	28-JUN-2000.
XX	
PF	20-DEC-1999; 99EP-0125406.
XX	
PR	23-DEC-1998; 98RU-0123511.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
PI	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
XX	
DR	WPI; 2000-414602/36.
DR	N-PSDB; A48443.
XX	
PT	Novel Escherichia bacterium having enhanced L-threonine resistance due
PT	to enhanced RhtC protein activity, used to produce L-threonine,
PT	L-homoserine, L-valine and L-leucine -
XX	

PS Claim 1; Page 15-16; 24pp; English.

XX The present sequence is the L-threonine resistance protein, RhtB, from
CC Escherichia coli. The coding sequence may be used to impart L-threonine
CC resistance on E. coli bacteria, which would be useful for producing
CC a high yield of L-threonine. L-threonine resistance means that the
CC bacteria will be able to grow on a minimal medium containing
CC L-threonine at a concentration at which the corresponding wild-type
CC strain would not grow. Since the transformed bacteria can grow on the
CC minimal medium, it can synthesize L-threonine, which accumulates. The
CC accumulated amino acids can then be removed from the culture medium.
CC The bacterium of the present invention may also be used to synthesize
CC L-homoserine, L-valine and L-leucine at increased levels.

SQ Sequence 206 AA;

Query Match 100.0%; Score 1054; DB 21; Length 206;

Best Local Similarity 100.0%; Pred. No. 1.9e-112;

Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLFLTVMVHIVAMSPGDPFFVSTAVSRSRKEAMMGVLGTCGMVWAGIALGL 60

|||||
Db 1 mmlfltvmvhiavmspgdpffvstavsrsrkeammgvlgtcgmvwagialgl 60

QY 61 HLIIEKMAWLHTLIMVGGGLYLCWMGYOMLRGALKKEAVSAPQVELAKSGRFLKGL 120

|||||
Db 61 hliiekmawlhtlimvgggllylcwmgyomlrgalkkeavsapqvelaksgsrlkgl 120

QY 121 TNLNPKAIIFGVSFSLFVGDNVGTTARWGIFALIIIVETLAWFTVVASLFPQMRG 180

|||||
Db 121 tnlanpkaiifgvsfslfvgdnvgttarwgifalilivetlawftvaslfpqmrgr 180

QY 181 QRLAKWIDGFAGALFAGFIHILISR 206

|||||
Db 181 qrlakwidgfagalafagfihliisr 206

RESULT 2

B01786

ID B01786 standard; Protein; 223 AA.

XX B01786;

DT 03-JAN-2001 (first entry)

DE Escherichia coli Yahn amino acid excretion protein.

XX E. coli; yahn gene; amino acid production; excretion protein gene;

KW amino acid excretion protein.

XX Escherichia coli.

OS EP1016710-A2.

PN 05-JUL-2000.

PD 17-DEC-1999; 99EP-0125263.

PF 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

XX (AJIN) AJINOMOTO CO INC.

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;

PI Tokhmakova IL;

XX WPI; 2000-414802/36.

DR N-PSDB; A52688.

XX Increased production of L-amino acids by an Escherichia bacterium

PT comprises increasing the expression amount of an L-amino acid excretion

PT protein -

XX Claim 1; Page 19; 29pp; English.

XX The present sequence is the Yahn amino acid excretion protein from
CC Escherichia coli. This protein is involved in the production of
CC amino acids, and an increase in its expression leads to an increase in
CC accumulation of amino acids in the cell. In this case, an increase in
CC lysine, glutamic acid and proline is achieved if multiple copies of its
CC gene are transfected into a bacterium. The bacterium used is E. coli.

SQ Sequence 223 AA;

Query Match 24.8%; Score 261.5; DB 21; Length 223;

Best Local Similarity 28.2%; Pred. No. 6.5e-22;

Matches 58; Conservative 50; Mismatches 83; Indels 15; Gaps 3;

QY 4 LFTVMVHIVAMSPGDPFFVSTAVSRSRKEAMMGVLGTCGMVWAGIALGLHLI 63

|||||
Db 20 vyltvl-fvitffnpganlfvvvgtslasgragvltglgvaldafysglglfglatl 78

QY 64 IEKMAWLHTLIMVGGGLYLCWMGYOMLRGALKKEAVSAPQVELAKSGRS-----FLK 117

|||||
Db 79 itceeiifslirivgagylwfwesmr-----rqstpqmstlqqpisapwyvffir 130

QY 118 GLTNLNPKAIIFGVSFSLFVGDNVGTTARWGIFALIIIVETLAWFTVVASLFPQMR 177

|||||
Db 131 glitdlnpqtvlffisfvtlnaetptwarlmawagivlasliiwrvfslsqafslpavr 190

QY 178 RGYORLAKWIDGFAGALFAGFIHIL 203

|||||
Db 191 raygrmqvrasvigaigvafalrli 216

RESULT 3

B01787

ID B01787 standard; Protein; 212 AA.

XX B01787;

DT 03-JAN-2001 (first entry)

DE Escherichia coli YeaS amino acid excretion protein.

XX E. coli; yeas gene; amino acid production; excretion protein gene;

KW amino acid excretion protein.

XX Escherichia coli.

OS EP1016710-A2.

PN 05-JUL-2000.

PD 17-DEC-1999; 99EP-0125263.

PF 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

XX (AJIN) AJINOMOTO CO INC.

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;

PI Tokhmakova IL;

XX WPI; 2000-414802/36.

DR N-PSDB; A52689.

XX Increased production of L-amino acids by an Escherichia bacterium

PT comprises increasing the expression amount of an L-amino acid excretion

PT protein -

XX Claim 1; Page 21; 29pp; English.

XX The present sequence is the Yeas amino acid excretion protein from

CC

CC Escherichia coli. This protein is involved in the production of amino
CC acids, and an increase in its expression leads to an increased
CC accumulation of amino acids in the cell. In this case, an increase in
CC lysine, alanine, valine, histidine, isoleucine, glutamic acid and proline
CC is achieved if multiple copies of its gene are transfected into a
CC bacterium. The bacterium used is E. coli.
XX
SQ Sequence 212 AA;

Query Match 18.0%; Score 190; DB 21; Length 212;
Best Local Similarity 28.2%; Pred. No. 8.8e-14;
Matches 57; Conservative 36; Mismatches 85; Indels 24; Gaps 6;

QY 13 IVALMSPGDPFFVSVOTAVSRKKEAMGVLTGCG-----VMVWAGTALLGLHLITEKM 67

Db 18 ifivlvpgpntlfvlnsvssgmkggyiaacgvfigdavlmlawagvatl-----lkt 72

QY 68 AWLHTLMVGGGLYLCWGMGYQMLRGALKKEAVSAPAPQVELAKSGRFLKGLTLNLANPK 127

Db 73 palfnivrlylgafilylsgkilyatlkgkseaksdepy---gaifkraliisltnpk 129

QY 128 AIIYFGSVSLFVGDNVGTARWGI FALIIVETLAW-----FTVVASLFPQMRRGY 180

Db 130 alifysffvqfidvnapht---gisffilaatlslvscylslfllisgaf-vtqyirtk 185

QY 181 ORLAKWIDGFAGALFAGFGIHL 202

Db 186 kkiakvgnseliglmfvfgfaarl 207

RESULT 4

ID B11637
XX B11637 standard; Protein; 229 AA.

AC B11637;
XX
DT 23-OCT-2000 (first entry)

XX A. vitis hypersensitive response elicitor protein, SEQ ID NO:9.

XX Hypersensitive response elicitor protein; HR; disease resistance;
KW insecticide; fungicide; antiviral; bactericide; growth enhancer;
KW stress resistance; transgenic plant.

XX Agrobacterium vitis.

XX WQ200028056-A2.

XX 18-MAY-2000.

XX 05-NOV-1999; 99WO-US26079.

XX 06-NOV-1998; 98US-0107387.

XX (CORR) CORNELL RES FOUND INC.

XX Burr TJ, Herlache TC, Zhang H;

XX WPI; 2000-376567/32.

XX N-PSDB; A61502.

XX New protein from Agrobacterium vitis, useful e.g. for imparting
PT resistance to disease or stress to plants, is involved in production of
PT a hypersensitive response

XX Claim 3; Page 85-86; 157pp; English.

XX Sequences B11630-B11688 represent proteins from Agrobacterium vitis which
CC elicit a hypersensitivity response (HR) in a plant. The invention also
CC relates to nucleotide sequences (A61501-A61524) encoding the A. vitis HR
CC elicitor proteins. The HR is a rapid, localised necrosis that is
CC associated with the active defence of plants against many pathogens, and

CC occurs when a pathogenic organism interacts with a nonhost plant (i.e.
CC one in which intracellular bacterial growth and disease development do
CC not occur). Like other HR elicitors, the A. vitis elicitor functions in
CC non-host plants by causing a rapid hypersensitive response that results
CC in walling-off and killing of the pathogen. On grape plants, the A. vitis
CC elicitor induces a restricted necrosis of tissues, resulting in the death
CC of plant cells and induction of pathogen resistance. A. vitis HR elicitor
CC proteins, in non-infectious form, are used to treat plants or their seeds
CC to impart resistance to disease, such as those caused by fungi, bacteria
CC or viruses; and to enhance growth, e.g., to increase yield or to provide
CC earlier germination or maturation. The proteins can also be used to
CC control insects, to impart resistance to environmental stresses, e.g.,
CC cold, and to improve nutritional value, e.g., altered oil content. The
CC same effects can be produced by producing transgenic plants or seeds by
CC incorporation of DNA that encodes A. vitis HR elicitor proteins. Use of
CC A. vitis HR elicitor proteins, or nucleic acids encoding them, may allow
CC control of previously untreatable diseases; provide systemic treatment;
CC and eliminate the need for biological control agents or polluting
CC chemicals.
XX
SQ Sequence 229 AA;

Query Match 17.0%; Score 179.5; DB 21; Length 229;
Best Local Similarity 27.0%; Pred. No. 1.5e-12;
Matches 54; Conservative 36; Mismatches 95; Indels 15; Gaps 5;

QY 6 LTVAMVHIVALMSPGDPFFVSVOTAVSRKKEAMGMVLGTGCMVMWAGTALLGLHLITE 65

Db 25 lliasitwlvslpgadfamvrsnflygrksglaaasmglaiacwfhviyafgialqh 84

QY 66 KMAWLHTLMVGGGLYLCWGMGYQMLRGALKK-----EAVSAPAPQVELAKSGRFLKGLT 121

Db 85 ifpnldilklfvgaylvvag---latafskirdvegslypsdr---smgremmtgilt 137

QY 122 NLANPKAIIYFGSVFLFVGDNVGTAR--WGIFALIIVETLAWFTTVASLFPQMRRG 179

Db 138 nglnpktsifvislytqfigkdtplshqllwglf--lslshllwfasvstflsnpairtv 195

QY 180 YORLAKWIDGFAGALFAGFG 199

Db 196 vlrrdrlfnlligvvlaslg 215

RESULT 5

Y99597
ID Y99597 standard; Protein; 205 AA.

XX Y99597;

XX 08-SEP-2000 (first entry)

XX E. coli L-homoserine resistance protein, RhtB.

XX L-homoserine resistance; L-homoserine synthesis; rhtB;
KW L-threonine; L-valine; L-leucine.

XX Escherichia coli.

XX EP1013765-A1.

XX 28-JUN-2000.

XX 20-DEC-1999; 99EP-0125406.

XX 23-DEC-1998; 98RU-0123511.

XX (AJIN) AJINOMOTO KK.

XX Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;

XX WPI; 2000-414602/36.

XX N-PSDB; A48442.

XX		Noel Escherichia bacterium having enhanced L-threonine resistance due
PT		to enhanced RhtC protein activity; used to produce L-threonine,
PT		L-homoserine, L-valine and L-leucine -
XX		
PS	Claim 2;	Page 13-14; 24pp; English.
CC	The present sequence is the L-homoserine resistance protein, RhtB, from	
CC	Escherichia coli. The coding sequence may be used to impart l-homoserine	
CC	resistance on E. coli bacteria, which would be useful for producing	
CC	a high yield of l-homoserine. L-homoserine resistance means that the	
CC	bacteria will be able to grow on a minimal medium containing	
CC	L-homoserine at a concentration at which the corresponding wild-type	
CC	strain would not grow. Since the transformed bacteria can grow on the	
CC	minimal medium, it can synthesise L-homoserine, which accumulates. The	
CC	accumulated amino acids can then be removed from the culture medium.	
CC	The bacterium of the present invention may also be used to synthesise	
CC	L-threonine, L-valine and L-leucine at increased levels.	
XX		
XX	Sequence	205 AA:
SQ		
	Query Match	11.6%; Score 122; DB 21; Length 205;
	Best Local Similarity	21.0%; Pred. No. 4.9e-06;
	Matches	46; Conservative 36; Mismatches 89; Indels 48; Gaps 8;
QY	10 MVHIVALSPGDPFFVSTAVSRKEMMGVLGITCGV-----MVGAGIALGL 60 : :: : :	:
Dg	11 ltsilflspgsaintmtstlnhgypag----gvycwasdrtdgscagwrgvgtlfs 65 : :: : :	:
QY	61 HLII--EKMAWLHTLIMVGGEYLTCMWGYQLR--GALKKEAVSAPAPOVELAKSGRSFL 116 : :: : :	:
Dg	66 rsvlafevikw-----agaayllwlqiqwraagaidlkxlastqtgrhl-----fq 112 : :: : :	:
QY	117 KGLLTNLANKAIIFGVSVFSFVDGNVGTTRARGIFALIIVEPLAWFTVVASLFALPOM 176 : :: :: : : :	:
Dg	113 rafvnaltnpkisvfialfpqlmpgpqlmgivlgvt-----tivvdilmvigy 164 : :: : :	:
QY	177 RRGYORLAKWIDG-----FAGALFAGFGIHLITSR 206 : :: : :	:
Dg	165 atlaiqrailwikpqmkalnifsglmvgallasar 203 : :: : :	:
RESULT	6	
ID	Y79298	
AC	X	standard; Protein; 205 AA.
AC	X	
AC	X	Y79298;
DT	T	
DT	T	18-JUL-2000 (first entry)
DE	E.	coli RhtB protein imparting homoserine resistance.
KW	Homo-	serine resistance; RhtB protein; L-homoserine; L-alanine;
KW	L-isoleucine;	L-valine; L-threonine.
OS	Esche-	rchia coli.
PN	EP994190-A2.	
PD	X	
PD	X	19-APR-2000..
PF	20-SEP-1999;	99EP-0118581.
PR	13-OCT-1998;	98RU-0118425.
PA	(AJIN)	AJINOMOTO CO INC.
PI	Livshits VA,	Zakataeva NP, Aleoshin VV, Belareova AV;
PI	ToKhmakova IL;	
DR	WIPI; 2000-273530/24 .	
DR	N-PSTD; Z94405.	

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
XX WPI; 2000-414802/36.
XX N-PSDB; A52690.
XX Increased production of L-amino acids by an *Escherichia* bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX Claim 1; Page 23; 29pp; English.
XX The present sequence is the yfik amino acid excretion protein from
CC *Escherichia coli*. This protein is involved in the production of amino
CC acids, and an increase in its expression leads to an increased
CC accumulation of amino acids in the cell. In this case, an increase in
CC threonine, histidine, glutamic acid and proline is achieved if multiple
CC copies of its gene are transfected into a bacterium. The bacterium used
CC is *E. coli*.
XX Sequence 195 AA;
SQ
Query Match 9.6%; Score 101; DB 21; Length 195;
Best Local Similarity 22.0%; Pred. No. 0.0012;
Matches 39; Conservative 32; Mismatches 72; Indels 34; Gaps 7;
QY 1 MMLFLFTVAWHI VALSPGDFPFVSOTAVSRKEAMGVLTGCG----VMVWAGIA 56
Db 5 llsafwtv---liamtppnllalsatsatshfrqstrvlagmslglvilmilcagis 61
QY 57 LGLHLHIEKMAWHLTLMVGGGLYLCVMGYQMLRGALKKEAVSAPAPQVELAKSGRSFL 116
Db 62 ----fslavidpaavh-lliswagaayivlwakiatsptkedglq-----kpsifw 108
QY 117 KGLTLNLANPKAIIFGVSFLFVGDNDVGTARWGIFALIIVETLAWFTVVASLFLAL 173
Db 109 asfalqfvnkvilly-----gvta-1stfvlpqatqalswvvgsvllam 151
RESULT 10
B40407
ID B40407 standard; Protein; 130 AA.
XX AC B40407;
XX DT 08-FEB-2001 (first entry)
XX DE Human ORFX ORF171 polypeptide sequence SEQ ID NO:342.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antineumatic; antihypertensive;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX OS Homo sapiens.
XX PN WO200058473-A2.
XX PD 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
XX N-PSDB; C74616.
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 608; 5507pp; English.
XX C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antihypertensive; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 130 AA;
Query Match 9.0%; Score 94.5; DB 21; Length 130;
Best Local Similarity 25.7%; Pred. No. 0.0038;
Matches 35; Conservative 22; Mismatches 58; Indels 21; Gaps 6;
QY 81 YLCVMGYQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLTLNLANPKAIIFGVSFLFV 140
Db 5 ylllyatwtwd--rsafamndtpvtatars--llirgflnlnpkltiffiaflpqfv 60
QY 141 GDNVGTTA-----RMGIFALIIVETLAWFTVVASLFLALPQMRGYY---QRLAKWIDGF 190
Db 61 --tpqgtapaqimvlsgvmfamtavflvlgllanvf----rravvesprvqnllr 113
QY 191 AGALFAGFGIHLIISR 206
Db 114 fatafaglglnlafaq 129
RESULT 11
R87527
ID R87527 standard; Protein; 153 AA.
XX AC R87527;
XX DT 01-JUL-1996 (first entry)
XX DE Mel-linked mlgA gene product.
XX Marine mela; selectable marker; oyster larva settlement;
KW pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
KW mlgA gene.

```

XX OS Shewanella colwelliana strain LSTDYF.
XX PN US5474933-A.
XX PD 12-DEC-1995.
XX PF 21-MAR-1990; 90US-0496804.
XX PR 08-NOV-1993; 93US-0148945.
XX PR 21-MAR-1990; 90US-0496804.
XX PR 10-NOV-1992; 92US-0974837.
XX PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX PI Fuqua WC, Weiner RM;
XX DR WPI; 1996-039515/04.
XX DR N-PSDB; T06767.
XX PT Novel gene encoding marine mela from Shewanella - useful as
XX PT selectable marker in genetic engineering and for inducing larval
XX PT oyster settlement
XX PS Example 7; Fig 14; 47pp; English.
XX CC The 5' end of an open reading frame (T06767) was identified
XX CC directly downstream of the Shewanella colwelliana mela gene
XX CC (T06766), and was designated mlgA (mel-linked gene). The
XX CC role of the encoded protein (R87527) was unclear as deletion
XX CC subcloning in E. coli demonstrated that only mela was required
XX CC for melanogenesis.
XX SQ Sequence 153 AA;

Query Match 8.6%; Score 91; DB 17; Length 153;
Best Local Similarity 25.5%; Pred. No. 0.012;
Matches 36; Conservative 31; Mismatches 54; Indels 20; Gaps 6;

QY 43 LGITCGV--MWV-----AGIAL-----LGLHLIERMAWLHTLMVGGGLYLCWMGYQ 88
DB 7 lgmsigvrtllwmvgelagvalvaiaavmgvasmmlnypqlfdilkwvgglylgylgis 66

QY 89 MLRGALKKEAVSAPQVELAKSGRSL-KGLLTNLNPKAIIYFGSVFLVGDNVGTT 147
DB 67 mwrakgkmanldntssqi---snralitqgfvtaianpkgwafmisllppfsisdqia 122

QY 148 ARWGFALIIVETLAWFTVVA 168
DB 123 pqimvllsimmt-effsmla 142

RESULT 12
W73358
ID W73358 standard; Protein; 153 AA.
XX AC W73358;
XX DT 11-FEB-1999 (first entry)
XX DE S. colwelliana MlgA protein sequence.
XX KW MlgA; marine bacterium; melanin synthesis; marine exopolysaccharide;
XX KW UV blocker; sunscreen; Mela.
XX OS Shewanella colwelliana.
XX PN US5846531-A.
XX PD 08-DEC-1998.
XX PF 07-JUN-1995; 95US-0476254.

```

```

XX PR 07-JUN-1995; 95US-0476254.
XX PR 21-MAR-1990; 90US-0496804.
XX PR 10-NOV-1992; 92US-0974837.
XX PR 08-NOV-1993; 93US-0148945.
XX PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX PI Fuqua WC, Weiner RM;
XX DR WPI; 1999-058995/05.
XX DR N-PSDB; V08553.
XX PT Mela melanin protein from marine bacteria - useful as UV blocker in
XX PT e.g. cosmetics
XX PS Example 7; Fig 13; 57pp; English.
XX CC This sequence is the Shewanella colwelliana MlgA protein, which is
XX CC related to the Mela protein of the invention. The invention relates to
XX CC Mela proteins from marine bacterium of the genus Shewanella, Vibrio or
XX CC Hyphomonas. The Mela protein is active in catalysing production of
XX CC melanin synthesis, as well as components of marine exopolysaccharides.
XX CC The Mela protein can be used as a UV blocker in sunscreen, pigment and
XX CC dye compositions useful in cosmetic applications, as tints for glass or
XX CC in paints and coatings for plastics, synthetic resins and fabrics, rubber
XX CC and wood.
XX SQ Sequence 153 AA;

Query Match 8.6%; Score 91; DB 20; Length 153;
Best Local Similarity 25.5%; Pred. No. 0.012;
Matches 36; Conservative 31; Mismatches 54; Indels 20; Gaps 6;

QY 43 LGITCGV--MWV-----AGIAL-----LGLHLIERMAWLHTLMVGGGLYLCWMGYQ 88
DB 7 lgmsigvrtllwmvgelagvalvaiaavmgvasmmlnypqlfdilkwvgglylgylgis 66

QY 89 MLRGALKKEAVSAPQVELAKSGRSL-KGLLTNLNPKAIIYFGSVFLVGDNVGTT 147
DB 67 mwrakgkmanldntssqi---snralitqgfvtaianpkgwafmisllppfsisdqia 122

QY 148 ARWGFALIIVETLAWFTVVA 168
DB 123 pqimvllsimmt-effsmla 142

RESULT 13
R98004
ID R98004 standard; Protein; 431 AA.
XX AC R98004;
XX DT 16-OCT-1996 (first entry)
XX DE Oestrogen-regulated pLIV1 gene-encoded protein.
XX KW Oestrogen-regulated gene; pLIV1; breast cancer; tumour; diagnosis;
XX KW marker; metastasis.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 125..138
XX FT /label= 1003
XX FT /note= "immunogenic peptide 1003, claim 11,
XX FT page 19"
XX FT Peptide 252..265
XX FT /label= 1004
XX FT /note= "immunogenic peptide 1004, claim 11
XX FT page 19"
XX FT Peptide 418..431

```



```
XX 09-DEC-1988; 88DE-3841454.
XX PA
XX (DEGS ) DEGUSSA AG.
XX Kassing F, Kalinowski J, Arnold W, Winterfeldt A, Puhler A;
XX Kautz P-S, Thierbach G;
XX WPI; 1990-186465/25.
XX DR NPSDB; Q04775.
XX
XX Site specific mutagenesis of DNA at restriction enzyme positions -
XX by treating corresp. restriction fragments with hydroxylamine,
XX and new mutated sequences and transformed microorganisms
XX
XX Disclosure; ; p; German.
XX
XX When expressed in host cells this confers Chloramphenicol resistance
XX on them.
XX
XX Sequence 390 AA;

Query Match 8.2%; Score 86; DB 11; Length 390;
Best Local Similarity 20.6%; Pred. NO. 0.14;
Matches 47; Conservative 35; Mismatches 78; Indels 68; Gaps 10;

QY 1 MLMFLTVAMVHIVALMSPGDPFFVSQ-----TAVSRSRKEAMMG 41
   ::: | | | | | | | | | | | | | | | | | | | | | |
Db 71 llivcllvfagshvigamtpvfslilitrvlsalanagflavalstattlvpankgrals 130
   :| | | | | | | | | | | | | | | | | | | | | |
QY 42 VL--GITCGVMVWAGI---ALLGLHLIIIEKMAWLHTLIMVGGGLYLCWMGYQMLRGA--- 93
   :| | | | | | | | | | | | | | | | | | | | | |
Db 131 illsgttiatvv--gvpagallgtalgwrttfwaiafcipaav-----gvirgvtnn 181
   : : | | | | | | | | | | | | | | | | | | | | |
QY 94 LKKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKAIIVFGSVSLFVGDNVGTTARMGIF 153
   : : | | | | | | | | | | | | | | | | | | | | |
Db 182 vgrsetsatsprirve-----lsqatprlil----amalgalinggtfaaftl 227
   | : | | | | | | | | | | | | | | | | | | | | |
QY 154 ALIIVETL----AWFTVVASLFALPQMRGYSQRLAKWIDGFAGALFAG 197
   | | | | | | | | | | | | | | | | | | | | | |
Db 228 apivtetaglaawvsvvalvmfg-----lgsflgvtiaq 261
   | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: May 6, 2001, 14:38:49
Job time: 395 sec

Db 67 MWRKGMANLNTSSQI-----SNRALITQGFVTAIANKPGWAFMISLLPPFISVDQAIA 122
QY 148 ARWGIFALIIVETLAWFTVVA 168
Db 123 POLMWLLSIIMMT-EFFSMLA 142

RESULT 2

5474933-4
; Patent No. 5474933
; APPLICANT: EINER, RONALD M.; FUQUA JR., WILLIAM C.
; TITLE OF INVENTION: MARINE MELA GENE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,945
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 974,837
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: 496,804
; FILING DATE: 21-MAR-1990
; SEQ ID NO: 4:
; LENGTH: 153
5474933-4

Query Match 8.6%; Score 91; DB 6; Length 153;
Best Local Similarity 25.5%; Pred. No. 0.0064;
Matches 36; Conservative 31; Mismatches 54; Indels 20; Gaps 6;

QY 43 LGYTGVV-MVW-----AGIAL-----LGLHLIEKMAWHLTLINVGGLYLCNMGYQ 88
Db 7 LGMSIGVRTLMMWGLAGVAIVAAMVGVASMLNYPQLFDLKWVGGLYLGIGIS 66
QY 89 MLRGALKKEAVSAPAPQVELAKSGRSFL-KGLLTNLANKPAIIYFGSVFSLFVGDNVGTT 147
Db 67 MWRKGMANLNTSSQI-----SNRALITQGFVTAIANKPGWAFMISLLPPFISVDQAIA 122
QY 148 ARWGIFALIIVETLAWFTVVA 168
Db 123 POLMWLLSIIMMT-EFFSMLA 142

RESULT 3

US-08-311-023-2
; Sequence 2, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; BREAST TUMOURS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-311-023-2

Query Match 8.2%; Score 86.5; DB 1; Length 431;
Best Local Similarity 27.4%; Pred. No. 0.085;
Matches 49; Conservative 21; Mismatches 56; Indels 53; Gaps 11;

QY 53 AGIALGLHLIEKMAWHLTLINVGGLYLCNMGYQMLRGALKKEAVS----- 101
Db 263 AGVATL-----AW---MWINGDGLHNFSDGLAI--GAFTGLSSGLSTSVAVFCH 308
QY 102 FAPQ-----VELAKSGRSFLKGLLTNLANKPAIIYFGSVFSLFVGDNVGTTARWGIFALI 156
Db 309 ELPHELGDFAVLLKAGMTVVKQAVLYN-ALSAMLAYLGMATGIFIGHYAENVSMW-IFAL- 365
QY 157 IVETLAWFTVVASLFPQM-----RRGYORLAKWIDGFAGALFAGFGIHLII 204
Db 366 ---TAGLFMYVALVDMVPEMLHNDASDHGCSRWGYFFLQN-----AGMLL-GFGIMLLI 415

RESULT 4

US-09-335-409-11
; Sequence 11, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-11

Query Match 7.8%; Score 82; DB 4; Length 713;
Best Local Similarity 22.0%; Pred. No. 0.55;
Matches 46; Conservative 29; Mismatches 76; Indels 58; Gaps 9;

QY 1 MMLFLFVAMVHIVALMSGPDPFFVTSQTAVSRSRKEAMGVL--GITCGVMWAGIAL 58
Db 9 VLLSLVTLALILVTAR-----ASGELARRLRQPEVLGELFGVGLGPSV-VGALAP 58
QY 59 GLH-----LIIEKMAWHLTLINVGGLYLCNMGYQMLRGALKKEAVSAPAPQVEL 108
Db 59 GFHRLAQEPAGVGVLSGISW-----IGALLILLMAGIEVDVGLRKEA----- 102
QY 109 AKSGRSFLKGLLTNLANKPAIIYFGSVFSLFVGDNVGTTARWGIFALIIVETLAWFTVVA 168
Db 103 ----RPGALSALGAIAAPLAA---GAAFSALVLDRLPS---GLFGLIVLSVATYSVIAK 152
QY 169 SLFALPQMRRGY-----QRLAKWI 187


```
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Tobin, Matthew B.
; TITLE OF INVENTION: Multiple Drug Resistance Gene of
; TITLE OF INVENTION: Aspergillus Fumigatus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center, DC1501
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,734B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Craig, Anne I.
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: X-9681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-354-9570
; TELEFAX: 617-354-4043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-612-734B-2

Query Match 7.3%; Score 76.5; DB 2; Length 1349;
Best Local Similarity 15.9%; Pred. No. 5.5;
Matches 37; Conservative 29; Mismatches 76; Indels 91; Gaps 5;

QY 3 MFLTVAMVHVALMSPGPPFFVQTAIVSRKKA----- 38
DB 826 LMFFVVGIAFISLUSINGTAFACISRLIRRAKSAFRSLRQDISFFPDREENSTGALTS 885
QY 39 -----MKGVLGTCGVWVWAGIALGLHLHLEKMAWLHLMV----- 76
DB 886 FLSTETKNLSGSGVTLGTIMTSTTLGAAMIITAIAGMKLALVCISVVPILLACGFLRF 945
QY 77 -----GGGLYLCHWGYQMLRGALKKEAVSAPAPQV-----E 107
DB 946 YMLAQFOQRKSAVESGASTAC-----EATSAITRTVASLTREQDVWGVYHDQ 992
QY 108 LAKSGRSLKGLLTN-----LANPKAIIYFGSVFLFVGDVNGVTTARWGIFALII 157
DB 993 LOKGRKSLISVLRSLLVASQALVFCVAGLGFVYGGTLLGHHEYSIFRFV 1045

RESULT 11
US-08-922-837-2
; Sequence 2, Application US/08922837
; Patent No. 5888770
; GENERAL INFORMATION:
; APPLICANT: Chalker,, Allison F.
; APPLICANT: Feliu,, Maria M.2.
; APPLICANT: Brown,, James R.
; APPLICANT: Bryant,, Alexander P.
; TITLE OF INVENTION: No. 5888770el SpoilIE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
```

```
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,837
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10077
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 783 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-922-837-2

Query Match 7.0%; Score 73.5; DB 2; Length 783;
Best Local Similarity 25.3%; Pred. No. 5.8;
Matches 49; Conservative 24; Mismatches 64; Indels 57; Gaps 12;

QY 35 RKEAM-----MKGVLGTCGVWV-----WAGIALGLHLHLEKMAW 69
DB 36 RKEAQRMLISGIAIILIFAAFKLGAAGTTLNLRLLVGLSLAYLAIFGLLIYLFEEK 95
QY 70 -----LHTLMVGGGLYLCHWGYQMLRGALKKEAVSAPAPQVELAKSG---RSFLKGL 119
DB 96 IRKQGLSGFETIFAGLLIFAYLVWKGDKSVLKGTMAGVVDLTGFTTTFAGGG 155
QY 120 LTNLANPKAIIYFGSVFLFVGDVNGVTTARWGIFALIIIVETL---AWFTV-VASLFPALPQ 175
DB 156 LIGVA-----LYIPTAF-LF--SNIGTYFIGSI--LILVGLLVSPWSYDIAEFTS--- 202
QY 176 MRGQYQRLAKWIDG 189
DB 203 --RGP---AKNWEG 211

RESULT 12
US-08-750-723A-2
; Sequence 2, Application US/08750723A
; Patent No. 5981219
; GENERAL INFORMATION:
; APPLICANT: Flugge, Ulif-Ingo
; APPLICANT: Weber, Andreas
; APPLICANT: Fischer, Karsten
; TITLE OF INVENTION: DNA MOLECULES WHICH CODE FOR A PLASTID
; TITLE OF INVENTION: 2-OXOGLUTARATE/MALATE TRANSLOCATOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., Esq c/o Fish & NEAVE
; STREET: 1251 Avenue Of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA: US/08/750,723A
FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 20 782.4
FILING DATE: 15-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr., Esq.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: GFB-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-750-723A-2

Query Match 6.9%; Score 73; DB 2; Length 569;
Best Local Similarity 20.8%; Pred. No. 4.3;
Matches 43; Conservative 24; Mismatches 56; Indels 84; Gaps 8;

QY 34 SRKEAMGV-LGITGVVMVAG-----TALLGLHLII-----EKMAW--- 69
DB 362 TRNESTMAVTLTGLVFGKLGVDVTAAILGLSLVLLITGVVTKCLAESVAMDYL 421
QY 70 -----LHTLIMVGGGLYLCWMGYOMLRGAL----- 94
DB 422 TWFAALIAMAGYLNKYGLITWFSENVVGVVGGGLSLW---QMSFGVLVLLYFYSHYFFAS 478
QY 95 -----KKEAVSAPQVELAKSGRSFLKGLLTNLNPKAIITVFGSVFLV 140
DB 479 GAAHIGAMPTAFLVSASALGTPPFLAAIVLSFLNLMGLLTHYIGSAPVFGANY---- 534
QY 141 GDNVGTARWGI-FALIIVETLANVTV 166
DB 535 ---VPLPQWVGFLISIVNLIILGV 558

RESULT 13
US-08-924-747-8
Sequence 8, Application US/08924747
Patent No. 6083570
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,747
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL TYPE: SOYBEAN
TISSUE TYPE: SOYBEAN
IMMEDIATE SOURCE:
CLONE: SE3.03B09
US-08-924-747-8

Query Match 6.7%; Score 71; DB 3; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6;

QY 84 WMGY-----OMLRGALKKEAVSAPQVELAKSGRSFLKGLLTNLNPKAIITVFGSVFS 137
DB 100 WFAYIDEKWFVTSLSVLVAEDDEAKKHPFEQAEGLERLEEVEFNKYSEGA--YFG---- 153
QY 138 LFVGDNVG-TTARWGI-FALIIVETLANVTWASLFAIPOM-RRGYORLAKWIDGFA 191
DB 154 ---GDSIGFIDIGGFSF-----LSWMRVIEEMSGRKLDEKKHPGLTQWAEFTA 199

RESULT 14
US-09-247-373B-8
Sequence 8, Application US/09247373B
Patent No. 6168954
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 234
TYPE: PRT
ORGANISM: SOYBEAN
US-09-247-373B-8

Query Match 6.7%; Score 71; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6;

QY 84 WMGY-----OMLRGALKKEAVSAPQVELAKSGRSFLKGLLTNLNPKAIITVFGSVFS 137
DB 100 WFAYIDEKWFVTSLSVLVAEDDEAKKHPFEQAEGLERLEEVEFNKYSEGA--YFG---- 153
QY 138 LFVGDNVG-TTARWGI-FALIIVETLANVTWASLFAIPOM-RRGYORLAKWIDGFA 191
DB 154 ---GDSIGFIDIGGFSF-----LSWMRVIEEMSGRKLDEKKHPGLTQWAEFTA 199

RESULT 15
US-09-296-715-8
Sequence 8, Application US/09296715

; Patent No. 6171839
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,715
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
; IMMEDIATE SOURCE:
; CLONE: SE3.03B09
; US-09-296-715-8

Query Match 6.78; Score 71; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6:

QY 84 WMGY-----QMLRGALKKEAVSAPQVELAKSGRSFLKGLLTNLANPKAIYFGSVFS 137
DB 100 WPAYIDEKWFSLRSVLVAEDDEAKKPHFEQAEGLERLEEVFNKYSEGKA--YFG--- 153

QY 138 LFVGDNVG-TTARWGIFALIIVETILANFTVVASLFALPQM-RRGYORLAKWIDGFA 191
DB 154 ---GDSIGFDIDFGGSF-----LSMMRVIEEMSGRKLDEKKHPLGTOWAETFA 199

Search completed: May 6, 2001, 14:39:24
Job time: 415 sec

OM of: US-09-466-935-4 to: GenEmbl:* out_format : pfs

Date: May 6, 2001 3:43 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=x10
-Q/cgn2_1/USF10.spool/US09466935/runat_03052001_075709_14798/app_query.fasta_1.529
-DB=GenEmbl -QFMT=fastap -Suffix=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEX=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=45 -DOCALIGN=200 -THRX_SCORE=pcr -THRX_MAX=100 -THRX_MIN=0
-ALIGN=15 -MODEL=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=200000000 -USER=US09466935 @CENL_1.5396 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-466-935-4
Query length: 206
Database: GenEmbl:*
Database sequences: 1283235
Database length: 1216004940
Search time (sec): 1745.950000

score_list:	Strd	Orig	zScore	Escore	Len	Documentation
gb_pat1:AX030177	+	1054.00	1898.96	1.5e-97	840	AX030177 Sequence 3 from Patent
gb_bal:AE005613	+	1054.00	1874.21	3.5e-96	10592	AE005613 Escherichia coli O15
gb_bal:AE000458	+	1027.00	1824.50	2.1e-93	11509	AE000458 Escherichia coli K12
gb_bal:BC00W85	+	998.00	1751.74	2.3e-89	91414	MF7049 E. coli genomic sequen
gb_bal:3.TGYSTM81	+	983.00	1724.08	8.1e-88	96086	AF233324 Salmonella typhimur
gb_hg16:AC073768	+	954.00	1661.13	2.6e-84	279589	AC073768 Mus musculus clone
gb_hg7:AC020970	+	747.00	1288.88	1.4e-63	212936	AC020970 Mus musculus clone
gb_hg7:AC020885	+	717.00	1232.29	2.0e-60	268234	AC020885 Mus musculus clone
gb_hg17:AC079167	+	559.00	945.36	1.9e-44	290452	AC079167 Mus musculus chromo
gb_bal:BC00REC0	+	485.00	858.85	1.3e-39	2695	M30198 E.coli recQ gene comple
gb_hg7:AC020893	+	411.00	692.35	2.4e-30	62274	AC020893 Mus musculus clone F
gb_bal:AE006110	+	370.50	636.68	3.0e-27	10184	AE006110 Pasteurella multocid
gb_bal:032810	+	368.50	632.19	5.4e-27	11137	U32810 Haemophilus influenzae
gb_bal:AE004109	+	313.50	532.41	1.9e-21	11321	AE004109 Vibrio cholerae chr
gb_bal:AP001517	+	292.00	461.66	1.7e-17	294250	AP001517 Bacillus halodurans
gb_pat1:AX063725	+	283.00	503.15	8.3e-20	792	AX063725 Sequence 7 from Patent
gb_bal:AE004719	+	274.50	462.85	1.5e-17	10145	AE004719 Pseudomonas aerugin
gb_bal:YEFYUABC	+	269.50	460.86	1.9e-17	4923	Z47200 Yersinia enterocolitica
gb_bal:AE004589	+	263.50	440.55	2.5e-16	12948	AE004589 Pseudomonas aerugin
gb_pat1:AX030079	+	261.50	465.81	9.9e-18	6724	AX030079 Sequence 9 from Patent
gb_bal:AE000140	+	261.50	437.27	3.9e-16	12498	AE000140 Escherichia coli K12
gb_bal:AE005212	+	258.50	433.98	5.9e-16	10029	AE005212 Escherichia coli O15
gb_bal:AP003009	+	255.00	393.24	1.1e-13	339681	AP003009 Mesorhizobium loti
gb_bal:AF235020	+	251.50	425.93	1.7e-15	6246	AF235020 Brucella melitensis s
gb_bal:BCU73857	+	251.50	396.37	7.3e-14	128824	U73857 Escherichia coli chrc
gb_bal:AE004699	+	246.50	411.87	1.0e-14	10426	AE004699 Pseudomonas aerugin
gb_bal:AE004864	+	245.00	408.65	1.5e-14	10976	AE004864 Pseudomonas aerugin
gb_bal:AF07482	+	244.00	398.33	5.7e-14	26245	AF07482 Pseudomonas aerugin
gb_hg8:AC022157	+	217.50	366.86	3.2e-12	4833	AC022157 Homo sapiens chromos
gb_bal:AF270347	+	214.50	367.20	3.1e-12	2675	AF270347 Staphylococcus epide
gb_bal:AF269817	+	214.50	366.31	3.5e-12	2329	AF269817 Staphylococcus epide
gb_bal:AE004946	+	208.50	340.84	9.1e-11	13075	AE004946 Pseudomonas aerugin
gb_bal:AE004889	+	207.00	339.45	1.1e-10	11415	AE004889 Pseudomonas aerugin
gb_bal:AP001508	+	201.00	296.76	2.6e-08	296950	AP001508 Bacillus halodurans
gb_bal:AP003001	+	199.50	292.55	4.4e-08	345783	AP003001 Mesorhizobium loti
gb_bal:PACARAB	+	193.00	339.86	1.0e-10	2483	U04992 Pseudomonas aeruginosa
gb_bal:PAU81259	+	193.00	329.34	4.0e-10	7285	U81259 Pseudomonas aeruginosa
gb_bal:AE004412	+	197.50	318.46	1.6e-09	16814	AE004412 Vibrio cholerae chr
gb_bal:AP002995	+	197.00	287.99	8.0e-08	346897	AP002995 Mesorhizobium loti
gb_pat1:AX030081	+	190.00	336.81	1.5e-10	639	AX030081 Sequence 11 from Patent
gb_bal:AE000274	+	190.00	306.81	7.1e-09	13793	AE000274 Escherichia coli K12

gb_bal:D90824 - 190.00 306.50 7.4e-09 14236 ! D90824 E.coli genomic DNA,
gb_bal:D90823 - 190.00 305.69 8.2e-09 15475 ! D90823 E.coli genomic DNA,
gb_bal:AF188935 + 190.00 287.84 8.1e-08 96231 ! AF188935 Bacillus anthraci
gb_bal:AP001519 + 186.00 269.37 8.7e-07 303650 ! AP001519 Bacillus halodur
seq_name: gb_pat1:AX030177

seq_documentation_block:

LOCUS AX030177 840 bp DNA PAT 16-SEP-2000
DEFINITION Sequence 3 from Patent EP1013765.
ACCESSION AX030177
VERSION AX030177.1 GI:10190394
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 840)
Balareva,A.V., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L. and
Zakataeva,N.P.
Gene and method for producing l-amino acids
Patent: EP 1013765-A 3 28-JUN-2000;
AJINOMOTO KK (JP)
FEATURES
source
1..840
/organism="Escherichia coli"
/db_xref="taxon:562"
187..807
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09319.1"
/db_xref="GI:10190395"

BASE COUNT

165 a 192 c 248 g 235 t
ORIGIN
alignment_scores:
Quality: 1054.00 Length: 206
Ratio: 5.117 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-466-935-4 x AX030177 ..
Align seg 1/1 to: AX030177 from: 1 to: 840

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
187 AGTGTGATGTTATTTCTCACCCTGCCATGGTGACATTGTGGCGCTTAT 236
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArg 34
237 GAGCCCGGTCGCCGATTCTCTTTTGTCTCTCAGACGCGTGTGAGT 286
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
287 CCGGTAAAGACGATGATGGCGTGTGGCATACCTCGCGCGTAATG 336
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisIleIleIleIle 67
337 GTTGTGGCGGGGATGGCGCTGTGGCTGCTGATTTGATTATCGAAAAAT 386
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuLeuCys 84
387 GGCCTGGCTGCATACGCTGATTATGTTGGCGGTGGCGTGTATCTCTGCT 436
84 rpMetGlyThrGlnMetLeuArgGlyAlaLeuLysGluAlaValSer 100

```
437 GGATGGCTTACCAGATCTACGTGGTGACCTGAAAGAGCGGTTTCT 486
101 AlaProAlaProGlnValGluLeuAlaIySSerGlyArgSerPheLeuIy 117
487 GCACCTCGCCACAGGTCGAGCTGGCGAAGATGGCGCAGTTTCTGAA 536
117 sGlyLeuLeuThrAsnLeuAlaasnProIySAlaIleIleIyPheGlys 134
537 AGGTTTACTGACCAATCTGCTAATCCGAAGCGATTATCTACTTTGGCT 586
134 erValPheSerLeuPheValGlyAspAsnValGlyThrAlaArgThrP 150
587 CGGTTCTCATTTGTCGCTGATACGTTGGCACTACCGCGCGCTGG 636
151 GlyTlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
637 GGCATTTTTCGCTGATCATCTGCAACGCTGGCGTGTATACCGTCGT 686
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
687 TGCAGCGCTTTGGCCCTGCGCAATGCGCGTGTATCAACGCTCTGG 736
184 lAlyStrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyTle 200
737 CGAAGTGGATTGATGTTTGGCGGCGCTTATTTGGCGGATTGGCATT 786
201 HisLeuIleIleSerArg 206
787 CATTTGATTATTTCGCGG 804

seq_name: gb_bai:AE005613

seq_documentation_block:
LOCUS AE005613 10592 bp DNA BCT 25-JAN-2001
DEFINITION Escherichia coli O157:H7 genome, contig 3 of 3, section 232 of 290.
ACCESSION AE005613 AE005174
VERSION AE005613.1 GI:12518685
KEYWORDS
SOURCE Escherichia coli O157:H7.
ORGANISM Escherichia coli O157:H7
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 10592)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,D.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
2 (bases 1 to 10592)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,D.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source
1..10592
Location/Qualifiers
/organism="Escherichia coli O157:H7"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:83334"
/note="enterohemorrhagic"
<1..71
misc_feature
/note="O-island #155; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(72..836)
/gene="z5332"

gene
complement(72..836)
/gene="z5332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to
residues 1 to 131 of 161 from Escherichia coli K-12 Strain
MG1655: B3815"
/codon_start=1
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKMITLNLKRIFLALTLPLFAVAADDCALSDPTLT
VOATVNPQTERVKWYQKANGAWGTFLHALLADMSQGVOMAMNGGTIDSYAPLG
LYLENGCOQVALNLASGEHFFLRPGGVYAGKVGIVRLDAFKTSKEIOFAVQSGP
MLMENSVINPRIHENVASRKIRNGVINGKNGNAYFLLSQATNFYDFACYAKALNVE
QLLYLDGTSIHMKMGGAIPQWIPYFVIMISVERKG"
1206..2156
/gene="corA"
/note="z5333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules:
Cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655: B3816"
/codon_start=1
/transl_table=1
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MLSAFOLENNRLTRLEVESQPLVNAVWIDLVPPDDDERLRVQS
ELGSLATRPSELEDIEASARFEDDGLHIHSFFFEDEADHAGNSTVAFTIRDCGLR
PIRELPAPFLYBMBARSQSMVDGNAYELLIDLETKLEQADETENIYSDLEQLSR
VIMESQGDGDEDALELDELDIGWVRLCLMDTORALNFLVRARLPGGLEQARE
LIRDLESLLPHNESLFOKVNFLMQAMGFNIEQNRIKIFSVSVVFLPFLVASSY
GMNFEPMLKSWFGYGAIFMILAGLAPLYFKRKNML"
2157..4848
/note="O-island #156; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="z5334"
complement(2198..2689)
/gene="z5334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILLCFLCCIYNLWLRPVLITYSDAGNSVFLVV
DHLPTDSDKINWYLNKQNEIKNOHPLPEGSWHTVWVIDINGFTDYKRYIGPYEDL
YCFPTIKSNDNCITKNYLMVINEYPYRTHIGINDTEYQLTQENKIEQVFNPHFK
DNF"
complement(2686..3534)
/gene="z5335"
complement(2686..3534)
/gene="z5335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNYLPSPINPCTLPETISLFDNYSAADMOYGDVVEODFLSLG
LSDISAKVDYRLKJHYHPPGGSINVAFASSGKIKNSRECTDILFAEMKLANFWS
FFGQYKTLIEDLIEHFHFGNGSNFSQQLNLSFHEKINQYNSPIRNIKEICNGIN
STPSTGYQLIQIKTLKLSRLNKFNDFNSGLISVHDISAQRIISLSLFQRYA
IGWSATIHFAQDHFGLDVTDIKNKTSKYRFRFRIFLQRHKDFAKFFFTFNPTIE
RIENYL"
3748..3867
/gene="z5336"
3748..3867
/gene="z5336"

misc_feature
/note="O-island #155; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(72..836)
/gene="z5332"

gene
complement(72..836)
/gene="z5332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to
residues 1 to 131 of 161 from Escherichia coli K-12 Strain
MG1655: B3815"
/codon_start=1
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKMITLNLKRIFLALTLPLFAVAADDCALSDPTLT
VOATVNPQTERVKWYQKANGAWGTFLHALLADMSQGVOMAMNGGTIDSYAPLG
LYLENGCOQVALNLASGEHFFLRPGGVYAGKVGIVRLDAFKTSKEIOFAVQSGP
MLMENSVINPRIHENVASRKIRNGVINGKNGNAYFLLSQATNFYDFACYAKALNVE
QLLYLDGTSIHMKMGGAIPQWIPYFVIMISVERKG"
1206..2156
/gene="corA"
/note="z5333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules:
Cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655: B3816"
/codon_start=1
/transl_table=1
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MLSAFOLENNRLTRLEVESQPLVNAVWIDLVPPDDDERLRVQS
ELGSLATRPSELEDIEASARFEDDGLHIHSFFFEDEADHAGNSTVAFTIRDCGLR
PIRELPAPFLYBMBARSQSMVDGNAYELLIDLETKLEQADETENIYSDLEQLSR
VIMESQGDGDEDALELDELDIGWVRLCLMDTORALNFLVRARLPGGLEQARE
LIRDLESLLPHNESLFOKVNFLMQAMGFNIEQNRIKIFSVSVVFLPFLVASSY
GMNFEPMLKSWFGYGAIFMILAGLAPLYFKRKNML"
2157..4848
/note="O-island #156; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="z5334"
complement(2198..2689)
/gene="z5334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILLCFLCCIYNLWLRPVLITYSDAGNSVFLVV
DHLPTDSDKINWYLNKQNEIKNOHPLPEGSWHTVWVIDINGFTDYKRYIGPYEDL
YCFPTIKSNDNCITKNYLMVINEYPYRTHIGINDTEYQLTQENKIEQVFNPHFK
DNF"
complement(2686..3534)
/gene="z5335"
complement(2686..3534)
/gene="z5335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=1
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNYLPSPINPCTLPETISLFDNYSAADMOYGDVVEODFLSLG
LSDISAKVDYRLKJHYHPPGGSINVAFASSGKIKNSRECTDILFAEMKLANFWS
FFGQYKTLIEDLIEHFHFGNGSNFSQQLNLSFHEKINQYNSPIRNIKEICNGIN
STPSTGYQLIQIKTLKLSRLNKFNDFNSGLISVHDISAQRIISLSLFQRYA
IGWSATIHFAQDHFGLDVTDIKNKTSKYRFRFRIFLQRHKDFAKFFFTFNPTIE
RIENYL"
3748..3867
/gene="z5336"
3748..3867
/gene="z5336"
```

```
/function="orf; Unknown function"
/Note="Residues 1 to 39 of 39 are 100.00 pct identical to
residues 278 to 316 of 316 from GenPept 118 :
gi|5960250|gb|AAFP33440.1| (AF233324) S. typhimurium
magnesium transport protein (CORA) (SP:P31138) [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="MNFEMPELKWSGYPGAIILMILAGUAPLYFKRKNWL"
/complement(3892..4278)
/gene="25337"
/complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59013.1"
/db_xref="GI:12518691"
/translation="MFKSIRNIFPSKLERLSRFQFIYILTSMISLITLNDINLAL
SFILVINTVYNINIGAKRIRDTCGKAIYFVLISLGLTLFSIAAIEGVRCESLSVQ
LNVAWILFTIMLSAIPSAKMSLSAH"
/gene="25339"
/4883..4794
/gene="25339"
/function="orf; Unknown function"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59014.1"
/db_xref="GI:12518692"
/translation="MDSIWEMVTFGSCWKNPKYSKSDVFKAWLLWAIITLVLFIL
SLTTLNLRSLGSLGFCQILGRYTMISLGSWKYMGVLGPPVTNIVLFLILFFKK
R"
complement(4841..5728)
/gene="25340"
/complement(4841..5728)
/gene="25340"
/function="orf; Unknown function"
/Note="Residues 4 to 295 of 295 are 92.12 pct identical to
residues 8 to 299 of 300 from Escherichia coli K-12 Strain
MGI655: B3819"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59015.1"
/db_xref="GI:12518693"
/translation="MDAKQTRQGVLLALAAFYFWGAPAYFKLIYVPADELTHRVL
NSFFWVWMSICRQNSYLKTLQTKIFMLAVSAVLGGNWLFIWAVNNHMLEA
SLGFNPINVLNIGMLIGERFRMQWLAVILACGLVQLWTFGSGPILALGAFS
FAPGLVRKKIAYEAQGMLETWMLLPVAAIYLAIDSDTSHMGPMSPMLNLLLIA
AGIVTVPLLCFTAAATRLRLSLTGFFQYIGPTLMFLAVITYGKPGADRMVITFI
WVALAIFYMDAIYTORRKS"
complement(5780..6265)
/gene="25341"
/complement(5780..6265)
/gene="25341"
/function="orf; Unknown function"
/Note="Residues 1 to 161 of 161 are 100.00 pct identical
to residues 1 to 161 of 161 from Escherichia coli K-12
Strain MGI655: B3820"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59016.1"
/db_xref="GI:12518694"
/translation="MKCKDMSAVLTAEQALKLVGEMFVHMPFNALGMELERYEKE
```

```
FAQLAFKNQPMVGNNAQSILHGGVITASALDYAAGLCVGVSTLIRHETISEDELQRRL
SRMTGIDLRVDYLRPGRXERFTATSSLLRAGNKVAVARVELHNEQLYIASATATYKV
G"
6412..7281
/gene="pIdA"
/Note="25342"
6412..7281
/gene="pIdA"
/function="enzyme; Central intermediary metabolism:
phosphorus compounds"
/Note="Residues 1 to 289 of 289 are 100.00 pct identical
to residues 1 to 289 of 289 from Escherichia coli K-12
Strain MGI655: B3821"
/codon_start=1
alignment_scores:
Quality: 1054.00 Length: 206
Ratio: 5.117 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-466-935-4 x AE005613
Align seg 1/1 to: AE005613 from: 1 to: 10592
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMet 17
9307 ATGTTGATGTTATTCTCACCGTCGCCATGTGCACATTGTGGCGCTTAT 9356
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
9357 GAGTCCCGGTCCTCCGATTTCTTTTGTCTCTCAGACCGCTGTCAGTCGT 9406
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
9407 CCCGTAAGAAGCGATGATGGCGTCTGGGTATTACCTGGCGGCTTATG 9456
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
9457 GTTTGGCAGGATGCGTTCGCTTGGCTTCGATTTGATTCGAAAAAAT 9506
67 tAlaTrpLeuHisThrIleMetValGlyGlyLeuThrLeuLeuCyst 84
9507 GCCTGGCTGCATACGCTGATGATGTTGGCGGTGGCTGATCTCTGCT 9556
84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
9557 GGATGGGTTACAGATGCTGCGCGCGCTGAAAAAAGAGCGCGTTCT 9606
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
9607 GCACCTGGCCACAGGTCGAGCTGGCGAAAAAGTGGCGCAGTTTCTGAA 9656
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIleThrPheGlyS 134
9657 AGGTTACTGACCAATCTCGCTAATCCGAAAGCGATATCTACTTTGGCT 9706
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
9707 CGGTGTTCTCATTTGTCGCTGATAACGTTGGCACTACCGCGCGCTGG 9756
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
9757 GGCATTTTCGGTGTGATATTATGTCGAAACGCTGGCTGTTTACTGTGGT 9806
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
9807 GCGCAGCTGTTTGGCTGCCGCAATGCGCGGTGGTTATCAACGCTCTGG 9856
184 lAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
9857 CGAAGTGATGATGTTTTCGCGGGCGGTATTTCGCGGATTTGGCATT 9906
```

```

201 HistLeuIleSerArg 206
|||||
9907 CATTGATTATTCGCGG 9924

seq_name: gb_ba1:AE000458

seq_documentation_block:
LOCUS AE000458 11509 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete
genome.
ACCESSION AE000458 U00096
VERSION AE000458.1 GI:2367299
KEYWORDS .
ORGANISM Escherichia coli K12.
          Escherichia coli K12.
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 11509)
          Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
          Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
          Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
          Mau, B., and Shao, F.
          The complete genome sequence of Escherichia coli K-12
          Science 277 (5331), 1453-1474 (1997)
97426617
REFERENCE 2 (bases 1 to 11509)
          Blattner, F.R.
          Direct Submission
          Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
          3 (bases 1 to 11509)
          Blattner, F.R.
          Direct Submission
          Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
          4 (bases 1 to 11509)
          Plunkett, G. III
          Direct Submission
          Submitted (13-OCT-1998) Laboratory of Genetics, University of
          Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
          On Sep 9, 1997 this sequence version replaced gi:1790254.
          This sequence was determined by the E. coli Genome Project at the
          University of Wisconsin-Madison (Frederick R. Blattner, director).
          Supported by NIH grants HG00301 and HG01428 (from the Human Genome
          Project and NCHGR). The entire sequence was independently
          determined from E. coli K12 strain MG1655. Predicted open reading
          frames were determined using GeneMark software, kindly supplied by
          Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
          30332 [email: mark@amber.gatech.edu]. Open reading frames that
          have been correlated with genetic loci are being annotated with CG
          Site Nos., unique ID nos. for the genes in the E. coli Genetic
          Stock Center (CGSC) database at Yale University, kindly supplied by
          Mary Berlyn. A public version of the database is accessible
          (http://cgsc.biology.yale.edu). Annotation of the genome is an
          ongoing task whose goal is to make the genome sequence more useful
          by correlating it with other data. Comments to the authors are
          appreciated. Updated information will be available at the E. coli
          Genome Project's World Wide Web site
          (http://www.genetics.wisc.edu). **The E. coli K12 sequence and
          its annotations are periodically updated; this is version M54. No
          sequence changes. Annotation updates: updated gene identifications
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.

Location/Qualifiers
1..11509
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
81..950
/gene="plda"
/note="b3821"
81..950
/gene="plda"
/EC_number="3.1.1.32"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/note="o289; 99 pct identical to PAL_ECOLI SW: P00631;
CGSCNo. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/transl_table="MRTLCQWLLPVMPLPMVYQAENVKEVHDPAVYRGSIANNMLQ
EHDNPFLLPYDITNYLIYTQSLNKEATASDMAENARKDEVKFQSLAPLRWGIL
GPNVIGASYTQKSWLSNSESSESPFETNYEPQLFGFATDYRFAGWTFDVMGTY
NHDNGRSQDPTSRSNRLYTRLMAENGMLVEYKPMYVVGNTDDDDITKTYGYQIK
IGYHLGDVILSAKGQYNWNTGYGGAGELGLSYPITKRVRLYTQVSYSGYESLIDYNFQ
TRVGVGVMINDLF"
965..993
/note="factor Sigma70; predicted +1 start at 4003392"
1019..1047
/note="factor Sigma70; predicted +1 start at 4003446"
1077..2909
/gene="recQ"
/note="b3822"
1077..2909
/gene="recQ"
/EC_number="3.6.1.-"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="o610; 99 pct identical to 607 amino acids
of RECQ_ECOLI SW: P15043 but has 3 additional N-ter
residues"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/transl_table="MNVAAQAEVLNLESGAKVQLQETFGYQVFRPQEEIIDTVLSGRD
CLVTPMGGKSLCYQIPALLNGLTVVYVSPILSMKQVQDQOLQANGVAAACLNSTQT
REQQLVYMTGTCGTQIRLYIAERLMDLNLEHLAHNPNVLLAVDEAHCSOMGHDF
RPEYALGQLORRFPPTLPFMALTATADDTTQDIDVRLCLNDPLLIQISSFORPNIRYM
LMERFKPLDOLMRVYQQRGSGIYYCNSRAKVEQDTEAARLQSGISAAAYHAGLENNV
RADYQEFQRDDLIIVVATVAFGNGINKPNRVFVHFDIPRNIESYQYQETGRAGRDGL
PAEAMLYEDADNMAVPRCLIEYPOGLODIERHKLNAMGAEQAQTCRRLLVLYRG
EGREPCGNCIDICLDPPKQDGSITQALISLTIGVNGQFNGCYVVEVIRGANNQRIE
DYHDKUKYVGMGRKDSHERWSVIRQLIHGLVTONIAQHSALQLTAEARPYLAESS
LQLAVPRIVALKPRAMQKRSFGNDRKLFKLFKLRKLSIADESNNVFPYVVFVNDATLIE
MAEQMPTIASEMLSYNGVGMKLERFKGPFMALIRAHVGDGDEE"
2104..2140
/note="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
3162..3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224..3592
/gene="yigU"
/note="b3823"
3224..3592
/gene="yigU"
/function="orf; Unknown"
/note="o122; sequence change shortens and
changesN-terminus relative to earlier version (YIG_ECOLI
SW: P27846)"
/codon_start=1
repeat_region
promoter
gene
CDS

```


JOURNAL Nucleic Acids Res. 21 (15), 3391-3398 (1993)
 MEDLINE 93347969
 REFERENCE 3 (bases 1 to 91414)
 AUTHORS Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and Daniels,D.L.
 TITLE Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes
 JOURNAL Nucleic Acids Res. 21 (23), 5408-5417 (1993)
 MEDLINE 94089392
 REFERENCE 4 (bases 1 to 91414)
 AUTHORS Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and Rouviere,P.E.
 TITLE A new family of peptidyl-prolyl isomerases
 JOURNAL Trends Biochem. Sci. 20 (1), 12-14 (1995)
 MEDLINE 95184296
 REFERENCE 5 (bases 1 to 91414)
 AUTHORS Daniels,D.L.
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 MEDLINE Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax: 608-263-7459
 REFERENCE On May 29, 1995 this sequence version replaced gi:148169.
 JOURNAL This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. The entire sequence was independently determined from E. coli MG1665. Overlaps and conflicts with other sequence determinations are annotated. Reference [1] describes the original sequence determination of a 91408 bp sequence. References [2], [3], and [4] describe subsequent corrections and/or updates to that sequence.

FEATURES Location/Qualifiers
 source 1..91414
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 /clone_lib="lambda library of Daniels and Blattner"
 /map="bp 1 at 3974 kb; 84.7 minutes"
 /note="This sequence comprises the following lambda clones: D850(EC15-116), D851(EC13RM4), D854(EC15-177), D856(EC19RM8.1), D858(EC17RM9), D864(EC21-76), D861(EC17-221), D864(EC27-149), D865(EC21-76), D867(EC27-860). M13mp19 or Janus vectors were used for subcloning."
 misc_feature 1..200
 /note="corresponds to K02846 (1..200)"
 misc_feature 1..5863
 /note="clone EC15-116"
 misc_feature complement(1..4899)
 /note="corresponds to M37337 (1..4900)"
 misc_feature 1..339
 /note="corresponds to V00349 (164..500)"
 tRNA 10..86
 /gene="asp"
 /note="CGSC No. 989"
 gene 10..86
 /gene="asp"
 tRNA 95..170
 /gene="trp"
 /note="CGSC No. 66"
 gene 95..170
 /gene="trp"
 terminator 177..205
 /note="terminator-like sequence"
 misc_structure 253..396
 /note="predicted bend of 81 degrees"
 gene complement(266..862)
 /gene="f198"
 CDS complement(266..862)
 /gene="f198"
 /codon_start=1
 /transl_table=11

misc_structure 586..724
 /note="predicted bend of 77 degrees"
 gene complement(703..1104)
 /gene="f133"
 CDS complement(703..1104)
 /gene="f133"
 /note="CGSC No. 18010; possible pssR gene"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA67567.1"
 /db_xref="GI:148171"
 /translation="MDTELLKTFLEVSRTRHFGRAAESLYLTQSAYSFRIQLENQLG VNLFRHRNIRLTAAAGEKLLPYAETLMTSTWQAARKRWRIPHDITSLFSLVPAPRCGNV CLISGWDACIKRMPIQAYSKRELPLNGSLM"
 complement(1136..1165)
 /note="promoter matrix score of 63"
 1151..1179
 /note="promoter matrix score of 57"
 1223..11636
 /gene="o137"
 1223..11636
 /gene="o137"
 /codon_start=1
 /transl_table=11
 /function="unknown"
 /protein_id="AAA67568.1"
 /db_xref="GI:148172"
 /translation="MAESFTTNRYFDNKHYPGFSRHGDFTIKEAQLLHRRHGYAFNE LDLGKREPTEEEKLFVAVCGEREPEVWMSKMYTRIKRPRFTLSAVNRRLK VLKTPITLITINKGEMPFYVSSFCRWMRSRIAR"
 1562..11586
 /gene="o137"
 /note="terminator-like sequence"
 complement(1585..3135)
 /gene="f516"
 complement(1585..3135)
 /gene="f516"
 /note="similar to Mg chelatase subunit of Rhodobacter capsulata"
 /codon_start=1
 /transl_table=11
 /function="unknown"
 /protein_id="AAA67569.1"
 /db_xref="GI:148173"
 /translation="MHIMVTGGILMSLIVHTRAAALGVNAPPTIVTVHLSKGLPGLTM VGLPETTKEARDRVRSAIINSYEPYPAKKITINLAPADLPKEGGYDLPILAIALAA SEQFANKLDEYELVGEALTGALRGVPGAISATKISGRKIIVAKDNEDEGLIN GEGCLTADHQAQVAFLEGKHALEXPKPTDAVRALOHDLSDVIGQEQKRGLEITAA GGHNLLIGPFGTKTMLASINGLLPDLNNEEALESAAITSLVNAESYQKQWRQPF RSPHSAITAMVGGGAIPTGFEISLAHNGVLELDELPEFERTLDALREPTESQIHL LSRTAKITYPAREQLVAAMNPSPTGHYQGNHNRCTPEQTLRYLNRLSGPFLDRFDLS LEIPLPPGILSKTVVPGESSATVKQVMAARERQFKRONKLNALDSEIPIQFCCKLE SEDAMLESTLIHLGLSIRAWQRLKLVARTIADQSDITIRQHLQEAHSVYRAIDRL THLOKLTN"
 3287..3433
 /note="predicted bend of 79 degrees"
 3297..3326
 /note="promoter matrix score of 74"
 3336..3676
 /note="corresponds to M11651 (1..341)"
 3336..3737
 /note="corresponds to V00289 (1..402)"
 3336..12791
 /note="corresponds to M10313 (1..9456)"
 3379..5721
 /note="V00290; EC1VX(1..2343)"

```
gene
3458..3556
/gene="ilvL"
CDS
3458..3556
/gene="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGEDA operon leader peptide"
/protein_id="AA67570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVISVVVLIIPPCGAALGRGA"
3696..4679
/gene="ilvG"
3696..4679
/gene="ilvG"
/feature="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetylhydroy acid synthase II, large subunit"
/protein_id="AA67571.1"
/db_xref="GI:148175"
/translation="MNGAQVYHVALRAQGVNTVFGPGGAIMPVYDALYDGVGEHLIC
RHOQANAAIGYARATKTKGVCIATSGPGATNLITGLADALLDSIPVVAITGVQSAP
FIGTDAFQEVFLTGLSLACKHSLFVQSLPEELPRIMAEAFVACSGRCPLVDIIPKD
IQLASGDIPEVFTVENVAFHAEVQARMLAKQKPMLYVGGVGMQAQVAPALRE
FLAATKMFATCTLKLGAIVEADYPIYILGMLGHGTKAANFAVQEGDILLIATVAGRDDR
VTGKLTNPAPHASVIHMDIDPAEMNKLQAHVALQGLDNLALLPALQQLNQ"
4675..5340
/gene="o221"
4675..5340
/gene="o221"
/feature="the mutation ilvO2096 joins this ORF to ilvG,
restoring function missing in wild-type"
/pseudo
/codon_start=1
/variation
replace(4677..4678,"tatg")

alignment_scores:
Quality: 998.00 Length: 206
Ratio: 4.892 Gaps: 0
Percent Similarity: 99.029 Percent Identity: 99.029

alignment_block:
US-09-466-935-4 x ECOW85 ..
Align seg 1/1 to: ECOW85 from: 1 to: 91414
1 MetLeuMetLeuPheLeuThrValAlaMetValHisLeuValAlaLeuMe 17
60863 AAGTGTGATGTTATTCACCGTCGCATGGTGACATTTGGCGCTTAT 60912
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
60913 GAGCCCGCGTCCCGATTCTCTTTTCTCTCAGACCGCTGTCTAGTCTT 60962
34 erArGlyGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
60963 CCCGTAAAGAGCGCATGATGGCGTGTGGCATTTACCTGGCGCGTATG 61012
51 valTrpAlaGlyIleAlaLeuGlyLeuHisLeuIleGluLysMe 67
61013 GTTTGGGCTGGGATGGCGTGTGGCGTGCATTTGATTATCGAAAAAT 61062
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrlLeuCyst 84
61063 GGCTGGCTGCATACGCTGATTATGGTGGCGGTGGCTG.TATCTCTGCT 61111
84 rpMetGlyTyrlGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
61112 GGATGGGTACCAGATCTACCTGGTGCACCTGAAAAAGAGCGGTTCT 61161
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
61162 GCACCTGGCCACAGTGCAGCTGGCGAAAGTGGCGCAGTTTCTCTGAA 61211
```

```
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIleTyrlPheGlyS 134
61212 AGT.TTACTGACCAATCTCGCTAATCCGAAAGCATTTACTTTGGCT 61260
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrrp 150
61261 CGGTGTTCTCATTTGTTCGGTGAATACGTTGGCATTACCGCGCTGG 61310
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
61311 GGCATTTTTCGCTGATTCATTTGCGAAACGCTGGCGTGGTTTACCGCTG 61360
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrlGlnArgLeuA 184
61361 TGCAGCGCTGTTTTCGCTGCGCAATGCGCGTGGTTATCAACGCTGG 61410
184 lAlaYsTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
61411 CGAAGTGATGATGTTTTCGCGCGCTTATTTGCCGATTGGCATTT 61460
201 HisLeuIleIleSerArg 206
61461 CATTTGATTATTTGCGCG 61478
seq_name: gb_ba3:STVSTMD1
seq_documentation_block:
LOCUS STVSTMD1 96086 bp DNA BCT 11-FEB-2000
DEFINITION Salmonella typhimurium fragment STMD1.
ACCESSION AF233324
VERSION AF233324.1 GI:6960215
KEYWORDS
SOURCE Salmonella typhimurium LT2.
ORGANISM Salmonella typhimurium LT2.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 96086)
AUTHORS Washington University Genome Sequencing Center.
TITLE The Salmonella typhimurium Genome Sequencing Project
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 96086)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: scifiton@watson.wustl.edu or
jsplie@watson.wustl.edu
```

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs
GTRIMMER (Salzberg, S., Delcher, A., Kasif, A. and White, O. (1998)
NAR 26,544-548), and Genemark (Lukashin, A.V. and Borodovsky, M.
(1998), NAR 26,1107-1115.

FEATURES
source
1..96086
/organism="Salmonella typhimurium LT2"
/strain="SGSC1412"

KLDLVDAMTLOGADPKVSDSDQSNQVERACPTGSCSCMFNTANSNNGITRALGLSQPGN
GSLATHADRKOLFNLAKRIVELTKRYEONDESALPRNIAKAAFERAWITLIDJAMG
GSTNTVHLHLLAAQAEIDFTMSDIDLKSRKPOLCKVAPSTQKRYHMEVDVHRAGGVLG
ILGELDRAGLLNRNKNVLTLPOTLEQYDITVQDEAVKRMFRAGPAGIRTTQAFS
QPCWDSLDDRAACGIRSLAYASKDGLAVLYNGFAENGCIKVTAGVDDSDILKFTG
PAKVYESQDDAEALIGKVVGVGVVIRYEGPKGPGMQEMLYPTSFLLKSMGLGKAC
ALITDGRFSGGTSGLSIGHVSPFAASGCTIALIEDGTIADIPNRSIOLQISEABIA
ARREAOEARGDKAWTPKPNRQKQVSFALRAYASLATSADKGAVRDKSKLGG"
9801..11345

alignment_scores:
Quality: 983.00 Length: 206
Ratio: 4.842 Gaps: 0
Percent Similarity: 98.544 Percent Identity: 90.777

alignment_block:

US-09-466-935-4 x STYSTMD1 ..
Align seg 1/1 to: STYSTMD1 from: 1 to: 96086

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
|||||
60428 ATGATGATGCTATTTCACCGTGGCAATGGTCACATCGTTCCGCTAAT 60477
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
|||||
60478 GAGCCCCGCGCCGACTTTTCTTCGTGTCTCAAACTGCTGCAGCGTT 60527
34 exArGlyGluAlaMetMetClyValLeuGlyIleThrCysGlyValMet 50
|||||
60528 CCGGTAAAGAGACGATGATGGCGCATCTGGGCGATCCCTCGCGCGTGG 60577
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisIleIleGlyLysMe 67
60578 GTATGGCGGGCGTGGCGCTCTTGGCTGCATCTCATCATCGAAAAAT 60627
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyLeuCysT 84
|||||
60628 GGCTTGGCTGCATAGCATATGTTAGTGGCGGCGCTGTACTGTGCT 60677
84 rPMeGlyTyTrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
60678 GGATGGGTATCAGATGCTGCGCGCGCTGAAAAACAGGATGCGGCG 60727
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
60728 GCGTCTCTCCGACATGGAATCGCGCAGCGGCGCGAGCTTCTCAA 60777
117 scGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyPheGlyS 134
60778 AGGGCTGTGACCAATCTGTCGAATCTTAAAGCGATATCTATTTGGTT 60827
134 exValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
60828 CCGTTTTTTCATTTTTCGCGCATAACGTCGCGCTGGCGGCGCGCTCG 60877
151 GlyIlePheAlaLeuIleValGluThrLeuAlaTrpPheThrValVa 167
60878 GGTATTTTCGGCTTAATCCCTCGAAACGCTGCCCTGGTTTACCGTAGT 60927
167 lAlaSerLeuPheAlaLeuProGlnMetArgGlyTyTrGlnArgLeuA 184
60928 CGCCAGCGCTGTTCGCGCTCGCGAAATGCGCGCTGGCTATCATCGCTCG 60977
184 lAlaTyTrPheAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
60978 CGAATGATGATGATGGCTTTCGCGCGCTCTGTTTGGCGCTTGTGATT 61027
201 HisLeuIleIleSerArg 206
61028 CACCTGATTATTCGCGC 61045
seq_name: gb_htg16:AC073768

seq_documentation_block:
LOCUS AC073768 279589 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-359N18, WORKING DRAFT SEQUENCE, 60
unordered pieces.
ACCESSION AC073768
VERSION AC073768.1 GI:8810385
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 279589)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 279589)
DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1875058
Center clone name: RPCI-23_359N18

Summary Statistics
Consensus quality: 230709 bases at least Q40
Consensus quality: 253950 bases at least Q30
Consensus quality: 258916 bases at least Q20
Estimated insert size: 216000; agarose-fp estimation
Estimated insert size: 273689; sum-of-contigs estimation
Quality coverage: 9.54 in Q20 bases; agarose-fp estimation
Quality coverage: 7.53 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1009: contig of 1009 bp in length
* 1109: gap of unknown length
* 1110: contig of 1023 bp in length
* 2132: gap of unknown length
* 2232: contig of 1025 bp in length
* 3257: gap of unknown length
* 3357: gap of unknown length
* 4550: contig of 1193 bp in length
* 4650: gap of unknown length
* 6414: contig of 1764 bp in length
* 6514: gap of unknown length
* 6515: contig of 1043 bp in length
* 7557: gap of unknown length
* 7657: contig of 1228 bp in length
* 8885: gap of unknown length
* 8986: contig of 1381 bp in length
* 10366: gap of unknown length
* 10467: contig of 1225 bp in length
* 11691: gap of unknown length
* 11792: contig of 1325 bp in length
* 13116: gap of unknown length
* 13216: contig of 1018 bp in length
* 14234: gap of unknown length
* 14335: contig of 1138 bp in length
* 15472: gap of unknown length
* 15573: contig of 1117 bp in length
* 16689: gap of unknown length
* 16789: contig of 1049 bp in length
* 17838: contig of 1049 bp in length

```
* 17839 17938: gap of unknown length
* 17939 18188: contig of 1250 bp in length
* 18189 18288: gap of unknown length
* 18289 18826: contig of 1538 bp in length
* 20827 20926: gap of unknown length
* 21984 21984: contig of 1058 bp in length
* 21985 22084: gap of unknown length
* 22085 23158: contig of 1074 bp in length
* 23159 23258: gap of unknown length
* 23259 24597: contig of 1339 bp in length
* 24598 24697: gap of unknown length
* 24698 25784: contig of 1087 bp in length
* 25785 25884: gap of unknown length
* 25885 26929: contig of 1045 bp in length
* 26930 27029: gap of unknown length
* 27030 28535: contig of 1506 bp in length
* 28536 28635: gap of unknown length
* 28636 30028: contig of 1393 bp in length
* 30029 30128: gap of unknown length
* 30129 31331: contig of 1203 bp in length
* 31332 31431: gap of unknown length
* 31432 32511: contig of 1080 bp in length
* 32512 32611: gap of unknown length
* 32612 33889: contig of 1278 bp in length
* 33890 33989: gap of unknown length
* 33990 35192: contig of 1203 bp in length
* 35193 35292: gap of unknown length
* 35293 36574: contig of 1282 bp in length
* 36575 36674: gap of unknown length
* 36675 37944: contig of 1270 bp in length
* 37945 38044: gap of unknown length
* 38045 39129: contig of 1085 bp in length
* 39130 39229: gap of unknown length
* 39230 40571: contig of 1342 bp in length
* 40572 40671: gap of unknown length
* 40672 42172: contig of 1501 bp in length
* 42173 42272: gap of unknown length
* 42273 44167: gap of unknown length
* 44168 44267: gap of unknown length
* 44268 45726: contig of 1459 bp in length
* 45727 45826: gap of unknown length
* 45827 47489: contig of 1663 bp in length
* 47490 47589: gap of unknown length
* 47590 48678: contig of 1289 bp in length
* 48679 48978: gap of unknown length
* 48979 50766: contig of 1788 bp in length
* 50767 50866: gap of unknown length
* 50867 54237: contig of 3371 bp in length
* 54238 54337: gap of unknown length
* 54338 56520: contig of 2183 bp in length
* 56521 56620: gap of unknown length
* 56621 58133: contig of 1513 bp in length
* 58134 58233: gap of unknown length
* 58234 59527: contig of 1294 bp in length
* 59528 59627: gap of unknown length
* 59628 63467: contig of 3840 bp in length
* 63468 63567: gap of unknown length
* 63568 65727: contig of 2160 bp in length
* 65728 65827: gap of unknown length
* 65828 68787: contig of 2960 bp in length
* 68788 72316: gap of unknown length
* 72317 72416: gap of unknown length
* 72417 76575: contig of 4159 bp in length
* 76576 76675: gap of unknown length
* 76676 82272: contig of 5597 bp in length
* 82273 82372: gap of unknown length
* 82373 86199: contig of 3827 bp in length
* 86200 86299: gap of unknown length
* 86300 91846: contig of 5547 bp in length
* 91847 91946: gap of unknown length
* 91947 96394: contig of 4448 bp in length
* 96395 96494: gap of unknown length

* 96495 102483: contig of 5989 bp in length
* 102484 102583: gap of unknown length
* 102584 110476: contig of 7893 bp in length
* 110477 110576: gap of unknown length
* 110577 119241: contig of 8665 bp in length
* 119242 119341: gap of unknown length
* 119342 130059: contig of 10718 bp in length
* 130060 130159: gap of unknown length
* 130160 142204: contig of 12045 bp in length
* 142205 142304: gap of unknown length
* 142305 157417: contig of 15113 bp in length
* 157418 157517: gap of unknown length
* 157518 185392: contig of 27875 bp in length
* 185393 185492: gap of unknown length
* 185493 212310: contig of 26817 bp in length
* 212310 212409: gap of unknown length
* 212410 249959: contig of 37550 bp in length
* 249960 250059: gap of unknown length
* 250060 279589: contig of 29530 bp in length.

FEATURES             Location/Qualifiers
     source            1..279589
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-359N18"
                     /cclone_lib="RPCI mouse BAC library 23"
BASE COUNT      69254 a 68603 c 67231 g 68544 t 5957 others
ORIGIN

alignment_scores:      Quality: 954.00      Length: 210
                       Ratio: 4.654        Gaps: 4
Percent Similarity: 97.619 Percent Identity: 95.714

alignment_block:
US-09-466-935-4 x AC073768/rev ..

Align seg 1/1 to reverse of: AC073768 from: 1 to: 279589

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
30830 ATGTTGATGATTTATTCACCGTCGCCATGTCACATTGTGGCGCTTAT 30781
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
30780 GAGCCCGGTCGCCATTTCTTTTGTCTCTCAGACCGCTGTCAGTCGTT 30731
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
30730 CCGGTAAAGAGCGATGATGGCGTCTGGGCATTACCTGCGGCGTAATG 30681
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
30680 GTTGGGCTGGGATTTGGCTGCTGGCTGCATTGATTATCGAANAAT 30631
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCysT 84
30630 GGCCTGGCTGCATACGCTGATTTATGTTGGCGGTGGCTGTATCTCTGCT 30581
84 tPmetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAla.Valse 100
30580 GGATGGGTTACCAGATGCTACGTGGTGCACCTGAAAAAGAGCGCGTTTC 30531
100 tAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuL 117
30530 TGCACCTCGCCACAGCTGAGCTGGCGAAGAGTGGCGCAGCTTCTCTGA 30481
117 ysgGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGly 133
30480 AAGGTTTACTGACCATCGTGAATCCGAAAGCGATATATCTACTTTGCG 30431
134 SerValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTr 150
|||||
```

```
30430 TCGGTGTTCTCATTTGTTGCGTGATACGTTGGCACTACCGCGCGTG 30381
150 pGlyIlePheAlaLeuIle.IleValGluThrLeuAlaTrpPheThrVal 166
|||||
30380 GGGCATTTTTCGCTGATCATTGTGAAACGCTGCGTGGTTTACCGTC 30331
167 Val.AlaserLeu.PheAlaLeuProGlnMetArgArgGlyTyrGlnArg 182
|||||
30330 GTTCGAGCTGCTGCTTTCCTCCGCAATGCGCGTGGTTATCA.CGT 30282
183 LeuAlaIysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheG1 199
|||||
30281 CTGCGGAAGTGGATTGATGTTTTCGCGGCGGCTATTTCGCGGATTGG 30232
199 YlleHisIleIleSerArg 206
|||||
30231 CATTCAATTGATTTCGCGG 30210
seq_name: gb_htg7:AC020970

seq_documentation_block:
LOCUS AC020970 212936 bp DNA HTG 10-FEB-2000
DEFINITION Mus musculus clone Rp23-252M21, WORKING DRAFT SEQUENCE, 144
unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 212936)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
REFERENCE 2 (bases 1 to 212936)
JOURNAL Direct Submission
TITLE Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
* 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
12223: contig of 1177 bp in length
13492: contig of 1269 bp in length
14669: contig of 1177 bp in length
15886: contig of 1217 bp in length
17089: contig of 1203 bp in length
18308: contig of 1219 bp in length
19518: contig of 1210 bp in length
20710: contig of 1192 bp in length
21932: contig of 1222 bp in length
23606: contig of 1674 bp in length
25328: contig of 1722 bp in length
26649: contig of 1321 bp in length
27849: contig of 1200 bp in length
29135: contig of 1286 bp in length
30700: contig of 1565 bp in length
31877: contig of 1177 bp in length
33065: contig of 1188 bp in length
34267: contig of 1202 bp in length
35474: contig of 1207 bp in length
36650: contig of 1176 bp in length
37859: contig of 1209 bp in length
39072: contig of 1213 bp in length
40406: contig of 1334 bp in length
41630: contig of 1224 bp in length
42864: contig of 1234 bp in length
44200: contig of 1336 bp in length
46087: contig of 1887 bp in length
48227: contig of 2140 bp in length
49518: contig of 1291 bp in length
50910: contig of 1392 bp in length
52420: contig of 1510 bp in length
53613: contig of 1193 bp in length
55004: contig of 1391 bp in length
56300: contig of 1296 bp in length
57505: contig of 1205 bp in length
58686: contig of 1181 bp in length
gap of unknown length
```

```
* 58687 60407: contig of 1721 bp in length
*      gap of unknown length
* 60408 61632: contig of 1225 bp in length
*      gap of unknown length
* 61633 62880: contig of 1248 bp in length
*      gap of unknown length
* 62881 64220: contig of 1340 bp in length
*      gap of unknown length
* 64221 65420: contig of 1200 bp in length
*      gap of unknown length
* 65421 66605: contig of 1185 bp in length
*      gap of unknown length
* 66606 68682: contig of 2077 bp in length
*      gap of unknown length
* 68683 70712: contig of 2030 bp in length
*      gap of unknown length
* 70713 71895: contig of 1183 bp in length
*      gap of unknown length
* 71896 73462: contig of 1567 bp in length
*      gap of unknown length
* 73463 74806: contig of 1344 bp in length
*      gap of unknown length
* 74807 76205: contig of 1399 bp in length
*      gap of unknown length
* 76206 77580: contig of 1375 bp in length
*      gap of unknown length
* 77581 79081: contig of 1501 bp in length
*      gap of unknown length
* 79082 80314: contig of 1233 bp in length
*      gap of unknown length
* 80315 81899: contig of 1585 bp in length
*      gap of unknown length
* 81900 83140: contig of 1241 bp in length
*      gap of unknown length
* 83141 84665: contig of 1525 bp in length
*      gap of unknown length
* 84666 85888: contig of 1223 bp in length
*      gap of unknown length
* 85889 87155: contig of 1267 bp in length
*      gap of unknown length
* 87156 88435: contig of 1280 bp in length
*      gap of unknown length
* 88436 89955: contig of 1520 bp in length
*      gap of unknown length
* 89956 91277: contig of 1322 bp in length
*      gap of unknown length
* 91278 92513: contig of 1236 bp in length
*      gap of unknown length
* 92514 93917: contig of 1404 bp in length
*      gap of unknown length
* 93918 95448: contig of 1531 bp in length
*      gap of unknown length
* 95449 96800: contig of 1352 bp in length
*      gap of unknown length
* 96801 97978: contig of 1178 bp in length
*      gap of unknown length
* 97979 99179: contig of 1201 bp in length
*      gap of unknown length
* 99180 100437: contig of 1258 bp in length
*      gap of unknown length
* 100438 101827: contig of 1390 bp in length
*      gap of unknown length
* 101828 103117: contig of 1290 bp in length
*      gap of unknown length
* 103118 104550: contig of 1433 bp in length
*      gap of unknown length
* 104551 105854: contig of 1304 bp in length
*      gap of unknown length
* 105855 107523: contig of 1669 bp in length
*      gap of unknown length
* 107524 109208: contig of 1685 bp in length
*      gap of unknown length
* 109209 110430: contig of 1222 bp in length
```

```
*      * 110431 111662: contig of 1232 bp in length
*
alignment_scores:
  Quality: 747.00      gap of unknown length: 206
  Ratio: 3.698        Gaps: 0
  Percent Similarity: 98.058      Percent Identity: 96.602

alignment_block:
US-09-466-935-4 x AC020970 ..
Align seg 1/1 to: AC020970 from: 1 to: 212936

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
174261 ATGTTGATGTTATTTCTCCACCGTCGCATGGTGCACATTTGGCGCTTAT 174310
17  tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
174311 GAGCCCGGTCGCCGATTTCCTT.TTTGTCTCTCAGACCGCTGTCAGTCGTT 174359
34  erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
174360 CCCGTAAGAAGCGATGAT.GGCGTCTGGGCATTACCTGGCGGCTAATG 174408
51  ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
174409 GTTTGGGCTGGGATTGGCT.CTTGGCCTGCATTGTGATTATCGAAAAAT 174457
67  tAlaTrpLeuHisThrLeuMetValGlyGlyGlyLeuYrLeuCyst 84
174458 GGCCTGCTGCATACGCTGAT.ATGCTGGCGGTGGCTGTATCTCTGCT 174506
84  rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
174507 GGATGGGTTACGAGATGCTAGT.GGTGCACCTGAAAAAGAGCGGTTCT 174555
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
174556 GCACCTGGCCACAGGTGCGA.CTGGCGAAAAAGTGGCGCAGTTTCCCTGAA 174604
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlys 134
174605 AGGTTTACTGACCAATCTCGT.AATCCGAAAGCGATTATCTATTGGCT 174653
134  erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
174654 CGGTGTTCTCATTTGTGCTG.TATACGTTGGCAGTACCGCGCGCTGG 174702
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
174703 GGCATTTTGGCGCTGATCAT.GTCGAAACGCTGGCGTGTATTACCGTCGT 174751
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
174752 TGCCAGCCTGTTTGGCCCTGGG.CAAAGCGCGCTGTTATCAACGCTCGG 174800
184 laLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
174801 CGAAGTGGATTGATGGTTTTC.CGGGCGGTATTATTCGCGGATTTGGCATT 174849
201 HisLeuIleIleSerArg 206
174850 CATTTGATTATTTCGCGG 174867

seq_name: gb_htg7:AC020885

seq_documentation_block:
LOCUS AC020885 268294 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RP23-464I1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020885
VERSION AC020885.2 GI:6980212
KEYWORDS HTG; HTGS_PHASE0.
```



```

* 44512 45780: gap of unknown length
* 45781 47202: contig of 1269 bp in length
* 47203 48647: contig of 1422 bp in length
* 48648 49652: contig of 1445 bp in length
* 49653 50485: contig of 1005 bp in length
* 50486 51527: gap of unknown length
* 51528 53119: contig of 833 bp in length
* 53120 53623: contig of 1042 bp in length
* 53624 54669: contig of 1592 bp in length
* 54670 55309: contig of 504 bp in length
* 55310 56382: contig of 504 bp in length
* 56383 56695: contig of 1046 bp in length
* 56696 57345: contig of 1060 bp in length
* 57346 57505: contig of 1403 bp in length
* 57506 58908: gap of unknown length
* 58909 59968: contig of 1060 bp in length
* 59969 61377: gap of unknown length
* 61378 62291: contig of 1409 bp in length
* 62292 63319: contig of 914 bp in length
* 63320 63818: gap of unknown length
* 63819 64823: contig of 1028 bp in length
* 64824 65269: gap of unknown length
* 65270 65904: contig of 499 bp in length
* 65905 66815: gap of unknown length
* 66816 6815: contig of 1005 bp in length
* 6816 6904: contig of 446 bp in length
* 6905 7004: contig of 635 bp in length
* 7005 7100: gap of unknown length
* 7101 7200: contig of 911 bp in length
* 7201 7300: gap of unknown length
* 7301 7400: gap of unknown length
* 7401 7500: gap of unknown length
* 7501 7600: gap of unknown length
* 7601 7700: gap of unknown length
* 7701 7800: gap of unknown length
* 7801 7900: gap of unknown length
* 7901 8000: gap of unknown length
* 8001 8100: gap of unknown length
* 8101 8200: gap of unknown length
* 8201 8300: gap of unknown length
* 8301 8400: gap of unknown length
* 8401 8500: gap of unknown length
* 8501 8600: gap of unknown length
* 8601 8700: gap of unknown length
* 8701 8800: gap of unknown length
* 8801 8900: gap of unknown length
* 8901 9000: gap of unknown length
* 9001 9100: gap of unknown length
* 9101 9200: gap of unknown length
* 9201 9300: gap of unknown length
* 9301 9400: gap of unknown length
* 9401 9500: gap of unknown length
* 9501 9600: gap of unknown length
* 9601 9700: gap of unknown length
* 9701 9800: gap of unknown length
* 9801 9900: gap of unknown length
* 9901 10000: gap of unknown length

```

```

11593 AAAGTTTACTGACCAATCTCGTATCCGAAGCGATTATCTACTTTGG 11544
133 ySerValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgT 150
11543 CTGGTGTCTCTCATTTGTCGTGATAGCTTGGCACTACCGCGGCT 11494
150 rpGlyIlePheAlaLeuIleValGluThrLeuAlaTrpPheThrVal 166
11493 GGGCATTTTGGCGTATCATTTGCGAAACGCTGGCGTGTACCGTC 11444
167 ValAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrGlnArgLe 183
11443 GTTGCAGGCTGTTTCCCTGCCGCAATGCGCGTGTGTATCAACGCT 11394
183 uAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyI 200
11393 GCGCAAGTGGATTGATGGTGTTCGCGGGCGCTTATTTGCCGGATTGGCA 11344
200 leHisLeuIleIleSerArg 206
11343 TTCAATTGATTATTTGCGCG 11324
seq_name: gb_htgl7:AC079167
seq_documentation_block:
LOCUS AC079167 290452 bp DNA HTG 30-AUG-2000
DEFINITION Mus musculus chromosome 6 clone RP23-226D23 strain C57BL6/J, ***
SEQUENCING IN PROGRESS ***, 196 unordered pieces.
ACCESSION AC079167
VERSION AC079167.2 GI:9945047
KEYWORDS HTG; HTGS_PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 290452)
AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
High Throughput Mouse Sequencing
Unpublished
2 (bases 1 to 290452)
Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
Gordon,M., Goltz,J.S. and Kucherlapati,R.
Direct Submission
Submitted (23-AUG-2000) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
On Aug 30, 2000 this sequence version replaced gi:9885883.
-----Genome Center
Albert Einstein College of Medicine Center
Code: AECOM
Web site:
http://sequence.aecom.yu.edu/cgi-
bin/ws.exe/mouseDB/mouseSeq/mouseseqtable.hts
Contact: jhan@sequence.aecom.yu.edu
-----Summary Statistics
Center project name: AEN
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big dye; 100%
*Consensus quality: 196511 at least Q20
*Consensus quality: 172308 at least Q30
*Consensus quality: 138790 at least Q40
Estimated insert size: agarose-FP - N/A
**Estimated insert size: 286552 - sum-of-contigs
Quality coverage: agarose-FP - N/A
Quality coverage: 3.3x sum-of-contigs - N/A
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 196 contigs. The true order of the pieces
* is not known and their order in this sequence record is

```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 1	852: contig of 852 bp in length	
* 853	872: gap of unknown length	
* 873	1753: contig of 881 bp in length	30181
* 1754	1773: gap of unknown length	31015
* 1774	2658: contig of 885 bp in length	31880
* 2659	2678: gap of unknown length	31900: gap of unknown length
* 2679	3514: contig of 836 bp in length	33160: contig of 1280 bp in length
* 3515	3534: gap of unknown length	33180: gap of unknown length
* 3535	4449: contig of 915 bp in length	34033: contig of 853 bp in length
* 4450	4469: gap of unknown length	34053: gap of unknown length
* 4470	5384: contig of 915 bp in length	34972: contig of 919 bp in length
* 5385	5404: gap of unknown length	34992: gap of unknown length
* 5405	6273: contig of 869 bp in length	35875: gap of unknown length
* 6274	6293: gap of unknown length	36731: contig of 836 bp in length
* 6294	7095: contig of 802 bp in length	36751: gap of unknown length
* 7096	7115: gap of unknown length	37639: contig of 888 bp in length
* 7116	7981: contig of 866 bp in length	37659: gap of unknown length
* 7982	8001: gap of unknown length	38499: contig of 840 bp in length
* 8002	8729: contig of 728 bp in length	38519: gap of unknown length
* 8730	8749: gap of unknown length	39348: contig of 829 bp in length
* 8750	9875: contig of 1126 bp in length	39368: gap of unknown length
* 9876	9895: gap of unknown length	40397: gap of unknown length
* 9896	10631: contig of 736 bp in length	41469: contig of 1072 bp in length
* 10632	10651: gap of unknown length	41489: gap of unknown length
* 10652	11577: contig of 926 bp in length	41890: contig of 401 bp in length
* 11578	11597: gap of unknown length	41910: gap of unknown length
* 12454	12454: contig of 857 bp in length	42795: contig of 885 bp in length
* 12455	12474: gap of unknown length	42815: gap of unknown length
* 12475	13334: contig of 860 bp in length	43744: contig of 929 bp in length
* 13335	13354: gap of unknown length	43764: gap of unknown length
* 14279	14279: contig of 925 bp in length	44638: contig of 864 bp in length
* 14280	14299: gap of unknown length	44648: gap of unknown length
* 14300	15245: contig of 946 bp in length	45958: contig of 1310 bp in length
* 15246	15265: gap of unknown length	45978: gap of unknown length
* 15266	16058: contig of 793 bp in length	46827: contig of 849 bp in length
* 16059	16078: gap of unknown length	46847: gap of unknown length
* 16079	16945: contig of 867 bp in length	47807: contig of 960 bp in length
* 16946	16965: gap of unknown length	47827: gap of unknown length
* 16966	17850: contig of 885 bp in length	48376: contig of 549 bp in length
* 17851	17870: gap of unknown length	48396: gap of unknown length
* 17871	18722: contig of 852 bp in length	49297: contig of 901 bp in length
* 18723	18742: gap of unknown length	49317: gap of unknown length
* 18743	19556: contig of 814 bp in length	50182: contig of 865 bp in length
* 19557	19576: gap of unknown length	50202: gap of unknown length
* 20817	20816: contig of 1240 bp in length	51054: contig of 852 bp in length
* 20837	20836: gap of unknown length	51074: gap of unknown length
* 21757	21756: contig of 920 bp in length	51373: contig of 299 bp in length
* 21757	21776: gap of unknown length	51393: gap of unknown length
* 22518	22517: contig of 741 bp in length	52053: contig of 660 bp in length
* 22538	23457: gap of unknown length	52073: gap of unknown length
* 23458	23477: gap of unknown length	52944: contig of 871 bp in length
* 23478	24289: contig of 812 bp in length	52964: gap of unknown length
* 24290	24309: gap of unknown length	53339: contig of 575 bp in length
* 24310	25249: contig of 940 bp in length	53559: gap of unknown length
* 25250	25269: gap of unknown length	54374: contig of 815 bp in length
* 25270	26135: contig of 866 bp in length	54394: gap of unknown length
* 26136	26135: gap of unknown length	55253: contig of 859 bp in length
* 26156	27048: contig of 893 bp in length	55273: gap of unknown length
* 27049	28068: gap of unknown length	56527: contig of 1254 bp in length
* 27069	28003: contig of 935 bp in length	56547: gap of unknown length
* 28004	28023: gap of unknown length	57376: contig of 829 bp in length
* 28024	28874: contig of 851 bp in length	57396: gap of unknown length
* 28875	28894: gap of unknown length	57473: contig of 77 bp in length
* 28895	29777: contig of 883 bp in length	57493: gap of unknown length
* 29778	29797: gap of unknown length	58337: contig of 844 bp in length
* 29798	30160: contig of 363 bp in length	58357: gap of unknown length
* 30161	30180: gap of unknown length	58444: contig of 87 bp in length
		58464: gap of unknown length
		59288: contig of 824 bp in length
		59308: gap of unknown length
		60499: contig of 1191 bp in length
		60519: gap of unknown length
		61859: contig of 1340 bp in length

* 61860 61879: gap of unknown length
* 61880 62752: contig of 873 bp in length
* 62753 62772: gap of unknown length
* 62773 63623: contig of 851 bp in length
* 63624 63643: gap of unknown length
* 63644 64694: contig of 1051 bp in length
* 64695 64714: gap of unknown length

alignment_scores:
Quality: 559.00 Length: 116
Ratio: 4.861 Gaps: 0
Percent Similarity: 99.138 Percent Identity: 99.138

alignment_block:

US-09-466-935-4 x AC079167 ..

Align seg 1/1 to: AC079167 from: 1 to: 290452

91 ArgGlyAlaLeuLysLysGluAlaValSerAlaProAlaProGlnVal101 107
104050 CGTGGTCACTGAAAAAGAGCG.GTTTCTGCACCTGCCACAGTCTGA 104098

107 uLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuA 124
104099 GCTGGCGAAAAGTGGCGCAGTTCTGAAAGTTTACTGACCAATCTCG 104148

124 laSenProLysAlaIleIleTyPheGlySerValPheSerLeuPheVal 140
104149 CTAATCGCAAGAGGATATCTATTGGCTCGGTCTTCATTTGTTGTC 104198

141 GlyAspAsnValGlyThrAlaArgTyrGlyIlePheAlaLeuIle11 157
104199 GGTGATAACGTGGCACTACCGCGCGCTGGGCAATTTTGGCGTATCAT 104248

157 eValGlyThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaLeuP 174
104249 TGTGCAACGCTGGCGGTTTACCGCTGTTGCCAGCTGTTTGGCCCTGC 104298

174 roGlnMetArgArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPhe 190
104299 CGCAAAATGCGCGTGTATATACAGTCTGGCGAAGTGGATGATGTTT 104348

191 AlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIleSerArg 206
104349 GCCGGGCGGTATTGTCGGGATTTGGCAATTCATTTGATTATTTCGCG 104396

seq_name: gb_ba2:ECORECQ

seq_documentation_block:
LOCUS ECORECQ 2695 bp DNA BCT 26-APR-1993
DEFINITION E.coli recQ gene complete cds, and pldA gene, 3' end.
ACCESSION M30198
VERSION M30198.1 GI:147557
KEYWORDS detergent-resistant phospholipase A; phospholipase; pldA gene; recQ gene.

SOURCE Escherichia coli (strain K-12) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 2695)
AUTHORS Irino,N., Nakayama,K. and Nakayama,H.
TITLE The recQ gene of Escherichia coli K12: Primary structure and evidence for SOS regulation
JOURNAL Mol. Gen. Genet. 205, 298-304 (1986)
MEDLINE 87115164

FEATURES
Location/Qualifiers
source 1..2695
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
1..352
/gene="pldA"
<1..352
gene
CDS

/gene="pldA"
/codon_start=2
/transl_table=11
/product="detergent-resistant phospholipase A"
/protein_id="AAA24516.1"
/db_xref="GI:147558"
/translation="SWNRLYTRLMAENGWLVEVKPWVYVGNDDDDPDITKMGYYOL
KIGYHLGDAVLSAKQYNNWTGYGGAELGLSYPTTKHVRLYTQVYSGYGESLIDYFN
QTRVGVGVMNLDF"
479..2311
/gene="recQ"
479..2311
/gene="recQ"
/codon_start=1
/transl_table=11
/protein_id="AAA24517.1"
/db_xref="GI:147559"
/translation="MNVAAEVLNLESGAKVLOETGYQOFRPGQBEIIDTVLSGRD
CLVYMTGGKSLCYOIPALLINGLVVVSPLISLMDQDQLOAQAAGAAACLNSTOT
REQLEVTGCTGQTIRLLYIAPERLMDNLEHLAHWNVLLAVDEAHCISQWHD
RPEYALGQLRQRPFTLPFMALTATADTTQDILVRLGLNDPLIQISSFDRENIRYM
LMEKFKPLDLMRYVQEQRGSGITTCNSRAKVEDTAAALQSKGISAAAYHAGLENNY
RADVQKFORDDLIQIWAQVAFGMGINKPNRVFVVFDPINPTIESYQETGRACRDGL
PAEAMLFYDPADMAWLRCLBEKPGQLODIERHKLNAMGAFAEATCRRLVLLNVEG
EGROECNCDCILDPKQYDSDAOIALSTIGRVNOREGMGVVVEVIRGANNQRI
DYGHDKLVYGMGRDKSHEHWVSITRILHGLVTONIAQHSALOLTEAARPVLAESS
LQAVPRIVALAPKAMQSGNDKRLFKLRKRSIADESNNPVPYVVFNDATLIE
MAEQMPITASEMLSVNGVMGRKLERFGFPFMAIIPAHVDGDDDE"
BASE COUNT 618 a 680 c 764 g 633 t
ORIGIN

alignment_scores:
Quality: 486.00 Length: 107
Ratio: 4.629 Gaps: 0
Percent Similarity: 98.131 Percent Identity: 98.131

alignment_block:

US-09-466-935-4 x ECORECQ ..

Align seg 1/1 to: ECORECQ from: 1 to: 2695

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
2375 ATGTTGATGTTATTTCTCACCCTGCCATGGTGACATTTGGCGCTTAT 2424

17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArg 34
2425 GAGCCCCGTCCTCGATTTCTTTTGTCTCTCAGACCGCTGTCTCAGTCGTT 2474

34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
2475 CCGGTAAAGACCGATGATGGCGCTGCTGGGCATTAC.TGCGGCGTAATG 2523

51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGlyLysMe 67
2524 GTTTGGCTGGGATTCGCTGCTGGCTGCAATTGATTATCGAAAAAAT 2573

67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCys 84
2574 GCCCTGGCTCGATACGCTG.ATTATGGTGGCGGCTGCGCTGATCTCTGCT 2622

84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
2623 GGATGGGTACCATGCTAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2672

101 AlaProAlaProGlnValGlu 107

2673 GCACCTGCCACAGGTCTGAG 2693

seq_name: gb_htg7:AC020833

seq_documentation_block:

LOCUS AC020833 62274 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RG-MBAC_40H2, LOW-PASS SEQUENCE SAMPLING.
AC020833
AC020833.1 GI:6686475
VERSION HTG; HTGS_PHASE0.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 62274)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 62274)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
296: contig of 296 bp in length
297 gap of unknown length
1049: contig of 753 bp in length
1050 gap of unknown length
1812: contig of 763 bp in length
1813 gap of unknown length
2420: contig of 608 bp in length
2421 gap of unknown length
2953: contig of 533 bp in length
2954 gap of unknown length
3645: contig of 692 bp in length
3646 gap of unknown length
4766: contig of 1121 bp in length
4767 gap of unknown length
5296: contig of 530 bp in length
5297 gap of unknown length
5584: contig of 288 bp in length
5585 gap of unknown length
6454: contig of 870 bp in length
6455 gap of unknown length
6853: contig of 399 bp in length
6854 gap of unknown length
7566: contig of 713 bp in length
7567 gap of unknown length
8253: contig of 687 bp in length
8254 gap of unknown length
9058: contig of 805 bp in length
9059 gap of unknown length
9678: contig of 620 bp in length
9679 gap of unknown length
10272: contig of 594 bp in length
10273 gap of unknown length
10890: contig of 618 bp in length
10891 gap of unknown length
11695: contig of 805 bp in length
11696 gap of unknown length
12677: contig of 982 bp in length
12678 gap of unknown length
13130: contig of 453 bp in length
13131 gap of unknown length
13878: contig of 748 bp in length
14026: contig of 148 bp in length
gap of unknown length
14027 contig of 843 bp in length
14869: gap of unknown length
15738: contig of 869 bp in length
15739 gap of unknown length
15426: contig of 688 bp in length
16427 gap of unknown length
17161: contig of 735 bp in length
17162 gap of unknown length
17938: contig of 777 bp in length
17939 gap of unknown length
18715: contig of 777 bp in length
18716 gap of unknown length
19493: contig of 778 bp in length
19494 gap of unknown length
20212: contig of 719 bp in length
20213 gap of unknown length
20780: contig of 568 bp in length
20781 gap of unknown length
21499: contig of 719 bp in length
21500 gap of unknown length
21570: contig of 71 bp in length
21571 gap of unknown length
22378: contig of 808 bp in length
22379 gap of unknown length
22940: contig of 562 bp in length
22941 gap of unknown length
23943: contig of 1003 bp in length
23944 gap of unknown length
24650: contig of 707 bp in length
24651 gap of unknown length
25686: contig of 1036 bp in length
25687 gap of unknown length
25820: contig of 134 bp in length
25821 gap of unknown length
26268: contig of 448 bp in length
26269 gap of unknown length
27152: contig of 884 bp in length
27153 gap of unknown length
28057: contig of 905 bp in length
28058 gap of unknown length
29002: contig of 945 bp in length
29003 gap of unknown length
29991: contig of 989 bp in length
29992 gap of unknown length
30780: contig of 789 bp in length
30781 gap of unknown length
31514: contig of 734 bp in length
31515 gap of unknown length
32152: contig of 638 bp in length
32153 gap of unknown length
32933: contig of 781 bp in length
32934 gap of unknown length
33042: contig of 109 bp in length
33043 gap of unknown length
34310: contig of 1268 bp in length
34311 gap of unknown length
34992: contig of 682 bp in length
34993 gap of unknown length
35821: contig of 829 bp in length
35822 gap of unknown length
37013: contig of 1192 bp in length
37014 gap of unknown length
38147: contig of 1134 bp in length
38148 gap of unknown length
38852: contig of 705 bp in length
38853 gap of unknown length
39840: contig of 988 bp in length
39841 gap of unknown length
40874: contig of 1034 bp in length
40875 gap of unknown length
42387: contig of 1513 bp in length
42388 gap of unknown length
43258: contig of 871 bp in length


```
gene
/db_xref="GI:12721022"
/translation="MMLNLIVVHFFGLITPGDPFFVSVSLAASNRNRNALCAVIGITL
GVFWALASILGLAIFNTVTVLQGLVMTLGGGYLAYLGYLMLKQQNVVPEPVEQEQ
YKQTSIKETTKGLLNLSNAKAVIYFASVMSLVNLVNTQWQWSALLIVLETFE
FYAIVSVFSQQAQFYSOYSRYIDHLSGVIFLLGGVLYMSGIGENTPLISK"
complement(3944..4429)
/gene="pgpA"
/notes="PM0728"
complement(3944..4429)
/gene="pgpA"
/codon_start=1
/transl_table=11
/product="PgpA"
/protein_id="AAK02812.1"
/db_xref="GI:12721023"
/translation="MLKTPLOQLKSNPIHFLALGFAGLIRPAGTWCMSVMGLMGW
GLQVNLGFSFLTALCFALGCVLCQKTAEDMGVHDGAIWDEIVGIFVLLSLPN
LSPFWCLTAFILFEFDLKPYPPIRYFDHKLSESGFGLMDDLLAALYAVIVFILRW
P"
complement(4439..5431)
/gene="thil"
/notes="PM0729"
complement(4439..5431)
/gene="thil"
/codon_start=1
/transl_table=11
/product="Thil"
/protein_id="AAK02813.1"
/db_xref="GI:12721024"
/translation="MSHGEPDILQVETASKRTARKDVILSGDDCAITELPONORLY
ITTPDMVTHFLPSMRADLAYKAVTNLSLDAAGAPAWVSALITLPOYDAQWLS
DESQSFVYLHYNDVILGDTTKGPLSVITTAHGLIPKGLGCRASASGVNWTYVS
TIGDSAGLSLLSQPLNSDHEFLIQRHLRTPRVLLGLELAVSSLANAALDSGD
FIADUDHITQSCQGVLELDKPLSSLVNVFVGSAETFAALSGEYELCTFQSPDS
NKAKLERALAYIGVDYTCVGVQRAISPQNKKRIKFORGNPDIDITQSGFDHFK"
complement(5448..5882)
/gene="nusb"
/notes="PM0730"
complement(5448..5882)
/gene="nusb"
/codon_start=1
/transl_table=11
/product="Nusb"
/protein_id="AAK02814.1"
/db_xref="GI:12721025"
/translation="MTSQKHEKKISPRRRARECAVQALYSWYVQNSPAAIELNPMAE
QDLKGVDTAYFRRLFRQTAENVDAVDNIMIPYLDREVSELDPIEKAILRLAYELKFE
LDVPYKVINGAIEVAKVFGAEDSHKYVNGVLDKAPVLSRK"
complement(5888..6361)
/gene="ribH"
/notes="PM0731"
complement(5888..6361)
/gene="ribH"
/codon_start=1
/transl_table=11
/product="RibH"
/protein_id="AAK02815.1"
/db_xref="GI:12721026"
/translation="MKVLEGAAPNAKIAVVIARFNSFTNESLLEGALDKRLGQV
KENTILVRVFGARPLVRLAESKKYDIVALGTVIRGGTAHFYVAGASSGLG
QVAMNADIPVAFGVLTENIEQAIERACTRAGNKAGAEALVALEVMVNLQAIDAA"
6575..7786
/gene="tyrp_1"
/notes="PM0732"
6575..7786
/gene="tyrp_1"
/codon_start=1
/transl_table=11
/product="Tyrp"
/protein_id="AAK02816.1"
/db_xref="GI:12721027"
/translation="MKNKTFGSALLVAGTTIGAGMLAMPITSMAEMGFVTYLLILFILM
```

```
GLLSYALLFVEVYQRAETKNAGIATLAEOYFGLGRILATLSLVVFMVYLAISAVTGG
GSLIAGVLPPLGEHATPIISVFTVIVGVFIVISGAVDMTLRFELFMKILVAEYFVL
LMLPLVNGDMLLAMPKDLFIISASPVFFTSFGFHVIIIPSNISYLDGNIRRLRIAI1
TGTAIPLVAYILWQLATHGVEPQAQFVQILNTDPTLNGLTATYQATESAIISHAMRL
FTTIALITSPFGVSLDFDCLYDLKARVKIKTRVSLGLLTFPLPLIFALFYPRGFVM
ALGAGOMFTFVGLVLPVGMARARKRPDLPYRVMGGLTLFGLAILGILIMVPEL
IKAGVLPVIG"
complement(7842..8504)
/gene="PM0733"
complement(7842..8504)
/gene="PM0733"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAK02817.1"
/db_xref="GI:12721028"
/translation="MSQHTKQOVLSAFHPRSAIRRYDPNKKVSEEDFHYILELGRLSP
SSVSGEPHFVAVQDROLREOLKAVSWGCTOLDASVLYFLAKKHARYDSPYPHDI
LAKGLPETMOKALAIYERFORDLOVLDNERTLFDWACKQTYIALANMWTGAAMIG
IDSCPIEGFDYCAVNNQLLSGTGAEDPNEGVSVAVTFGYRAKEVPKSRKPLTEITTW
LK"
8727..10106
/gene="htrA"
/notes="PM0734"
8727..10106
/gene="htrA"
/codon_start=1
/transl_table=11
/product="htrA"
/protein_id="AAK02818.1"
/db_xref="GI:12721029"
/translation="MKRERLLTSIALGLSILSTPMISOASLPTQVDQGVPSLAPML
EKVLPAVVISVEGKARKNHGSTPQGIPEFRFGPDLFNDRAPNRFRGIGSVIIN
ADSKLRLVNNHVIDSADKIITVQLDQRELSAKVIGTDESDVALIOIEKPNLTAITI
ADSKLRLVNNHVIDSADKIITVQLDQRELSAKVIGTDESDVALIOIEKPNLTAITI
SGGPLINQGLIGINTAIISPSGNAGTAFALPSNMANNVQOILEFGEVRRGLMGI
KGGELNADLAKAFDIEAQGAFFSEVLPNSAEAKAGLKAQDVIVAMNGOKISSFAEMR
AKIATSGAKREIATYLRDGKTHQTKVTQSDQTDQASHLLPALAGAEMSNHDDKG
KGVLTITSVTSLAEQRLKKGDIITGVNRQKVENLQLRKILDAKPSATALNIVRG
NTNFYLLIQ"
BASE COUNT 3070 a 2201 c 1937 g 2976 t
ORIGIN

alignment_scores:
Quality: 370.50 Length: 205
Ratio: 2.406 Gaps: 2
Percent Similarity: 75.122 Percent Identity: 37.073

alignment_block:
US-09-466-935-4 x AE006110/rev ..

Align seg 1/1 to reverse of: AE006110 from: 1 to: 10184

4 LeuphLeuthrValAlaMetValHisIleValAlaLeuMetSerProgl 20
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3938 GTGATGTTAAATCTTATTATTGACATTTTTTCGGTTTAATCACGCCAGG 3899

20 yProaspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3888 ACCGACATTTTTTATGTCAGTCAGTAGGTCGCCAGTAATTCACGTCGTA 3839

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3838 ATGCACITTTGCGCGCTGATTGTTACCTTAGGCGTGTGTTTGGGCA 3789

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLe 70
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3788 TTAGCCTCTATTATTAGGTTTAGCCATATATTATTATACGGTTTCCGCTTT 3739

70 whisThrLeuIleMetValGlyGlyLeuTyrlLeuTyrlMetGlyT 87
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3738 ACAGGGATTAGTCATGACTTTAGGCGGTGGTATTTAGCCTACTTGGGTT 3689
```



```

LDVPKVVINEAIEVAKVFGADESHKVGINGVLDKIAPLGRK"
2625..3662
/ gene="H11305"
CDS
2625..3662
/ gene="H11305"
/ note="similar to GB:U00096 SP:P77785 PID:1773101
PID:1786619 PID:2329840 percent identity: 53.09;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="thiamin-monophosphate kinase (thiL)"
/protein_id="AAC22952.1"
/db_xref="GI:1574764"
/translation="MARVFNACLFYKIGKMGAMGEFDLIKRYFOOQIILVDSQVLSI
GDCALNVPENYQAIATIDWENTHFLPTISPEDLAYKANATNLSDLAAGAPQPKW
ICRLAULPNVDENITFSOSLHLLTKOYNVTIGGDTKGNLSITITAGGFVKNKG
ICRKAQIGDLYVSSLSAGLTQIILGKSAVSDVDVFLQORHLRPTPRIELGQA
LIGIAHVAIDSLGLISDLGHILERSQCSAEVELTALPISSLSLNKRYDRTOAEQFALS
GGDEYELCFTIPPEYKDELELRLLKNVPCTCIGKINEKCGDFSPFLRDGKPVNITF
SSGFQHFKESEK"
3659..4150
/ gene="H11306"
CDS
3659..4150
/ gene="H11306"
/ note="similar to GB:U00096 PID:1773102 PID:1786620
PID:2329841 percent identity: 47.77; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="phosphatidyglycerophosphatase A (pgpA)"
/protein_id="AAC22953.1"
/db_xref="GI:1574765"
/translation="MTENNPLKKISLINPHLLAVGFGSLIHPADCTWCSLAGTILG
VILLSLVKVKFLFTALCLFELGCLTKTADMGVHDGSIWDFGVGVFLAAIP
SLSWQILAAALFALREFEDILKPFPIRYFDEKLENGFEMIDDLVLAAYAVIVVFAIQ
WML"
4147..4779
/ gene="H11307"
CDS
4147..4779
/ gene="H11307"
/ note="similar to GB:U00096 SP:P75693 PID:1657525
PID:1786522 percent identity: 26.87; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22954.1"
/db_xref="GI:1574766"
/translation="MMLNLIIVHLGLMTPDPFFVYSRMAASNRNTVCGILGITL
GIAFGMISMLGLAVLFTPALHGVIMLIGSYLAYLGLFMARSKYAKFBSHSDTE
FNOQTIKEKILGLLYNLSNAKYVYFSSVMSLVLVNITEMQIILAPYIVVETFC
YFYVLSLFSRNIAKRLYSQYSRIIDNAGIVLFFGCVLVINGINEIHH"
4796..5608
/ gene="H11308"
CDS
4796..5608
/ gene="H11308"
/ note="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710
PID:216458 percent identity: 69.74; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="dihydrodipicolinate reductase (dapB)"
/protein_id="AAC22955.1"
/db_xref="GI:1574767"
/translation="MTLKIAIAGAGRMGQCQLIQVHSRAGVELGAARFKGSSLVGT
DAGELAGHIGVAVSDLESQDKFDLLIDFTRPEGTLEHAFVCVANNKRKVIITG
FDENGKVAIKAAASKIAIVFASNE SVGNLVFKLLEKAARVMGDYCDIEVIAHRHK
VDAPSGTAKSMGEHIAKTGLRDLKTHGVFCRIGITGERKDEIGFSTIRASDVGEHS
VWFADIGRVEISHKASRMTFANGAVRACKWLENKANGLFONTVDLNLNL"
5609..5851
/ gene="H11309"
CDS
5609..5851
/ gene="H11309"
/ note="similar to SP:P37910 GB:U00096 PID:1788568 percent
identity: 54.41; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAC22956.1"
/db_xref="GI:1574768"
/translation="MKIHLRHNTTLENNETSLDHLKNNITHHEYQCRSGYCGSCR
VKIKGKYSYKEMPLAFIQPDEILLCCHVESDIEIDL"
5876..6637
/ gene="H11310"
CDS
5876..6637
/ gene="H11310"
/ note="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1

alignment_scores:
Quality: 368.50 Length: 207
Ratio: 2.457 Gaps: 4
Percent Similarity: 72.464 Percent Identity: 38.647

alignment_block:
US-09-466-935-4 x U32810 ..
Align seg 1/1 to: U32810 from: 1 to: 11137

3 MetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerPr 19
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4144 GTTGATGATGCTAAATTTAATCATTTGTCATTTATTTGGATTATGACGCC 4193
19 OGlyProaspPhePhePheValSerGlnThrAlaValSerArgSerArgL 36
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4194 AGGGCTGATTTCTTTATTAAGTCGAATGGCGGCAAGTAAGTCTCGTC 4243
36 ysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrp 52
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4244 GTAATACAGCTTTGTGGCATTTTAGGCATACACGCTGGCATCGCTTTTGG 4293
53 AlaGlyIleAlaLeuGlyLeuHisLeuIleIleGluLysMetAlaTr 69
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4294 GAATAGCTTTCTATGTTGGATGGCGGTGGTGGTTGTTACCATTCACGC 4343
69 PheuHisThrLeuMetValGlyGlyGlyLeuTyrrLeuTyrrMetG 86
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4344 ATTACATGCGGTTATTATGTTGCTAGGTGGTAGTTACCTAGCATATCTCG 4393
86 lyTyrrGluMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaPro 102
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4394 GTTTTTTATGGCTCGCAGT.....AAAAAATACGCTAAATTTGAATCG 4437
103 AlaProGlnValGluLeuAlaLysSer.....GlyArgSerPheLe 116
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4438 CACTCTGACTAGTAAATTAATCAACAAACCAACCAATCAAAAAGAAATTTT 4487
116 uLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrrPheG 133
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4488 GAAAGGGCTTTTAGTGAATTTATCCATGCAAAAGTCGTGCTGATTATTA 4537
133 lyservValPheSerLeuPheValGlyaspasnValGlyThrThrAlaArg 149
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4538 GTAGCGTATGATCGCTGCTCTTA.....GTAATATACATCACTGAAATG 4578
150 TrpGlyIle.....PheAlaLeuIleIleValGluThrLeuAlaTr 163
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4579 TGGCAAAATATCTTGGCTTTTGCAGTGATTGTGGTGAACAACATTTGTTA 4628
163 pPheThrValValAlaSerLeuPheAlaLeuProGlnMetArgArgGlyT 180
||||| :||||| : :||||| : :||||| : :||||| : :||||| :
4629 TTTTATGTGATTCATGATTTTTCACGTAATATATGCGACGCTTTAT 4678
180 yrGlnArgLeuAlaLysTrpIleaspGlyPheAlaGlyAlaLeuPheAla 196

```



```
CDS
complement(7742..8410)
/genes="VC0191"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF93367.1"
/db_xref="GI:9654596"
/translation="WNMSILATAGVHFIALLSPGPDVALVQVONATQHRKRGVMTA
IGLSGILVHLILUSGISLVKQOPMLFNLIAGGSYLLYGAGALQVMAQKNAS
TPHSPAFSLQKQAFKQMTNLPKALVFFVLSLLSPASMSVSGKSAAI
LVGLSLTWFSCIAWLTLTTSAMQRMORITRSVDSICAFAVLGAVILWQASRALAQ
T
FWL"
complement(8485..9297)
/genes="VC0192"
complement(8485..9297)
/genes="VC0192"
/notes="similar to GP:3294250; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, AraC/XyIs family"
/protein_id="AAF93368.1"
/db_xref="GI:9654597"
/translation="WYSMDTVNRPSPALSLTEADYRQFAPRHHLDIHGLITQ
GVRFHQGAMHQVGGVWLMSPDELHDGHAHNTGYQVQVFSIEPEWQIQLEANQ
IEQVIGFEQLIVQDPTLPHSLQQLHLLRDNLSQAKDCLPYQGFAPIRLERYSYLQ
PVPKPLGQHNLALEKWLSQLDQPIRLEQLAQICQLSPTQFQRHFRQAQGLTPYAWL
RRLRLPQAMKLLQSGKPGTDVAYQVGFYDQAHFSAKAFKATYGLSPSLITRFS"
complement(9334..9429)
/genes="VC0193"
complement(9334..9429)
/genes="VC0193"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF93369.1"
/db_xref="GI:9654598"
/translation="WMGRVCPSEFTRTSFFPTRRGTDHFLQDKL"
complement(9445..11211)
/genes="VC0194"
complement(9445..11211)
/genes="VC0194"
/notes="similar to GB:AL123456; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="gamma-glutamyltranspeptidase"
/protein_id="AAF93370.1"
/db_xref="GI:9654599"
/translation="MEMGNMNTHTPLALSAMFAAAQATPNQATDAVAPEQATGFEH
KSLVAKNMVTAAPLASEAGASILRQGNADIAVMTQLMLGLVPEPSSGIGSGSF
LVYWDAAKKALTFDGRETAPLNAPELFDSTQPMKFDVAVVGGRSVGTPTGTVKLL
WETHRYGKLEWALIEPVAKLAEOGFVSPRLAALAEKRLGRVPAFKAYFFDAQ
GPELTAGTLKPNDAATLRAIAQOGASAFYQGDIAKDILATVONAPNGVLAQODF
DIYQVKQAPCAAYQSYQVCGMGIPSSSGLTVCGLALAEQYDLKQWGAQDVKSQV
IGDASOLAFADGLYMAQDYPVPTQGLLTKTYLAERAKLIQPKKALTAPAGNPPW
HRAQLSPQSLSTSHFNVDREGNVVFTSIEANAFGRSLLVKRGFLNNELTDF
STATQSEGRINARPEKRPNSMPTIVIODNOPYLAIGSPGGSRIIGYVQAAIVA
HTQWGMIOQAATNQPHVLRNFRFEIEQGTSAQFPALESIGAKVGKIKELNSGLHAI
RITAQGLEGAADPREGVAIGE"
BASE COUNT 2707 a 2798 c 3044 g 2772 t
ORIGIN

alignment_scores:
  Quality: 313.50 Length: 198
  Ratio: 2.118 Gaps: 2
Percent Similarity: 74.747 Percent Identity: 31.313

alignment_block:
us-09-466-935-4 x AE004109/rev ..
```

```
Align seg 1/1 to reverse of: AE004109 from: 1 to: 11321
2 LeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSe 18
::: : : : : : : : : : : : : : : : : : : : : : : : : : :
8401 ATGTCATTTTATAGCCACCTTAGCCGGTGTGCATTTTATTCGCTTAAAG 8352

18 rProGlyProAspPhePheValSerGlnThrAlaValSerArgSerA 35
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8351 CCGGGGCCCTTAGTGCCTTAGTGTGCAAAACGCTACACGACGCGGC 8302

35 rGlySGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMetVal 51
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8301 GTAAACGTGCGTGAATGCGGTAGGTTTGTGTCGGCATTTTAGTC 8252

52 TrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleLeuGlyLysMetAl 68
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8251 CATTGTGATTTTGAGCCTGTCCGGCATAGCTACTTAGTCAAAACGACGCC 8202

68 aTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuTyrTrpM 85
||| : : : : : : : : : : : : : : : : : : : : : : : :
8201 AATGCTGTTTAACTGTTTAACTGGCGGCGGACGCTACTTGCTTTAT 8152

85 eTGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer... 100
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8151 TAGGGCGCGCGCTTTGCAATCAGTATGCGCAAAACGCTAGCAC 8102

101 .....AlaProAlaProGlnValGluLeuAlaLysSerGlyArgse 114
::: : : : : : : : : : : : : : : : : : : : : : : :
8101 CCGACGCACAGCCCTGCGCCCTCCATT...CTCGGCAACCGCGCTCAAGC 8055

114 rPheLeuGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleLeu 131
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8054 CTTCACAAAAGGATGACCAATCTCTTGAATCCCAAGCTCTCGGT 8005

131 yRPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThrThr 147
||||| : : : : : : : : : : : : : : : : : : : : : : : :
8004 TTTTGTGACGCTGTATAGCTGATCCACGACGATGCTGTGTCT 7955

148 AlaArgTrpGlyIlePheAlaLeuIleValGluThrLeuAlaTrpPh 164
::: : : : : : : : : : : : : : : : : : : : : : : :
7954 GCGAGTGAGTCCGCTCGGATTTGGTTGGCTTATCCCTAATGTT 7905

164 eThrValValAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrG 181
::: : : : : : : : : : : : : : : : : : : : : : : :
7904 CTCCTGTTTATGCTGTGCTTACCACGCTCTGCCATGCAACACGATGC 7855

181 lnArgLeuAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPhe 195
||||| : : : : : : : : : : : : : : : : : : : : : : : :
7854 AGCGCATAGCGATCGTCGATAGCATATGCGCTGCGGTGTT 7811

seq_name: gb_ba2:AP001517

seq_documentation_block:
LOCUS AP001517 294250 bp DNA BCT 10-JAN-2001
DEFINITION Bacillus halodurans genomic DNA, section 11/14.
ACCESSION AP001517 BA000004
VERSION AP001517.1 GI:10175500
KEYWORDS
SOURCE Bacillus halodurans DNA.
ORGANISM Bacillus halodurans
Bacteria; Firmicutes; Bacillus/Clostridium group;
Bacillus/Staphylococcus group; Bacillus.
REFERENCE 1 (sites)
AUTHORS Takami,H., Nakasone,K., Hirama,C., Takaki,Y., Masui,N., Fujii,F.,
Nakamura,Y. and Inoue,A.
TITLE An improved physical and genetic map of the genome of alkaliphilic
Bacillus sp. C-125
JOURNAL Extremophiles 3 (1), 21-28 (1999)
MEDLINE 99184645
REFERENCE 2 (sites)
AUTHORS Takami,H., Nakasone,K., Ogasawara,N., Hirama,C., Nakamura,Y.,
Masui,N., Fujii,F., Takaki,Y., Inoue,A. and Horikoshi,K.
TITLE Sequencing of three lambda clones from the genome of alkaliphilic
```



```
gene
5508..5807
/ gene="BH2884"
CDS
5508..5807
/ gene="BH2884"
/ note="BH2884"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06603.1"
/db_xref="GI:10175506"
/translation="MSQNERDFFDLMFGRPPSTDAENPOESTSGSTVEETKTE
GETKQETNAAPALPLTFEQMEHIFRLAQSLGALKSLSPYVKTIQQLISQKKD"
complement(5903..6157)
/ gene="BH2885"
complement(5903..6157)
/ gene="BH2885"
/ note="BH2885"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06604.1"
/db_xref="GI:10175507"
/translation="MNYTHGYYRQYIGTGERFYPAVLPLFAGLAVGPLLFGGKP
IGGYGPAIGSYGNFNGPNFGPMYGPSPQPPGYR"
complement(6226..6555)
/ gene="BH2886"
complement(6226..6555)
/ gene="BH2886"
/ note="BH2886"
unknown
/codon_start=1
/transl_table=11
/protein_id="BAB06605.1"
/db_xref="GI:10175508"

alignment_scores:
Quality: 292.00 Length: 207
Ratio: 2.147 Gaps: 2
Percent Similarity: 65.700 Percent Identity: 32.367

alignment_block:
US-09-466-935-4 x AP001517 ..

Align seg 1/1 to: AP001517 from: 1 to: 294250

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProG1 20
:::||||::: ||| :::|||| :::||||:::|||| :::||||:::||||
52063 ATGTCGTGAGGATTCATCGTCGGCTGTTAGCAGGAATGCTCCTGG 52112
|||||||||:::|||| :::||||:::||||:::|||| :::||||
20 yProAspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
:::||||:::|||| :::||||:::||||:::|||| :::||||
52113 CCCAGACTTCTTATCGTCATCAAAACAGCCTAGGTTTGGAGCGAG 52162
|||||||||:::|||| :::||||:::||||:::|||| :::||||
37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
:::||||:::|||| :::||||:::||||:::|||| :::||||
52163 TTGGTATCCTAACTTCTTTAGGAGTTGCTTCCCTAATCGTTCATATT 52212
|||||||||:::|||| :::||||:::||||:::|||| :::||||
54 GlyIleAlaLeuGlyLeuHisIleLeuGluLysMetAlaTrpLe 70
:::||||:::|||| :::||||:::||||:::|||| :::||||
52213 ACATATACAGTACGATTTGCTTTTAAATTGAACATATCCTGCTCT 52262
|||||||||:::|||| :::||||:::||||:::|||| :::||||
70 uHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCysTrpMetGlyT 87
|||||||||:::|||| :::||||:::||||:::|||| :::||||
52263 TTTTTTTACGATTCAACTATTAGGAGCTGCTTATTAAATTGGTTAGTT 52312
|||||||||:::|||| :::||||:::||||:::|||| :::||||
87 yrGlnMetLeuArgGlyAla.....LeuLysLysGluAlaValSerAla 101
:::||||:::|||| :::||||:::||||:::|||| :::||||
52313 TTCATGCCATACGCTCATCTCCCTCAAAAAGGAGAGCTGAAATCGAA 52362
|||||||||:::|||| :::||||:::||||:::|||| :::||||
102 ProAlaProGlnValGluLeuAlaLysSerGlyArgSer..... 114
:::||||:::|||| :::||||:::||||:::|||| :::||||
52363 GAAACACACCAATCCATCAACAAAGATAGTAAATCATCCATCAAGG 52412
```

```
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
|||||
187 ATGTTGATGTTATTTCTACCGTCGCATGGTGCACATTTGTGGCGTTAT 236
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
|||||
237 GAGCCCGGTCGCCGATTTCTTTTGTCTCAGACCGCTGTGAGTCGTT 286
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
|||||
287 CCGCTAAGAAGCGATGATGGCGTCTGGGCATTACCTGCGGCGTAATG 336
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMe 67
|||||
337 GTTTGGCTGGGATGGCTGCTGGCTGGCTGCTGCTGCTGCTGCTGCT 386
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCys 84
|||||
387 GGCCTGGCTGCATACGCTGATTATGTTGGCGGTGGCTGTATCTCTGCT 436
84 rPheMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
|||||
437 GGATGGGTTACAGATGCTACGTGGTGCACCTGAAAAAGAGCGGTTCT 486
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeu 117
|||||
487 GCACCTGCCGCACAGTCGAGCTGGCGAAAGTGGCGCAGTTCTCTGAA 536
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGly 134
|||||
537 AGGTTTACTGACCAATCTCGCTAATCGAAAGCGATTATCTACTTTGGCT 586
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTyr 150
|||||
587 CGGTGTTCTCATTTGTCGGTGATAACGTTGGCACCTACCGCGCGCTGG 636
151 GlyIlePheAlaLeuIleValGluThrLeuAlaTrpPheThrValVa 167
|||||
637 GGCATTTTGGCGTGATCATGTGCGAACGCTGGCGTGGTTTACCGTCGT 686
167 lAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrGlnArgLeuA 184
|||||
687 TGCAGCGCTGTTTGCCTGCGCGAAATGCGCGCTGTTATCAACGCTCG 736
184 lAlaSerTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGly 200
|||||
737 CGAAGTGGATTGATGGTTTGGCCGGGCGGTATTTCGCCGATTTCGCATT 786
201 HisLeuIleIleSerArg 206
|||||
787 CATTTGATTATTCGCGG 804
```

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52688

```
seq_documentation_block:
ID A52688 standard; DNA; 672 BP.
XX
AC A52688;
XX
XX 03-JAN-2001 (first entry)
DT
DE Escherichia coli yahn gene.
DE
KW E. coli; yahn gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
XX Key Location/Qualifiers
FH 1..672
FT /*tag= a
FT /product= "yahn"
XX
```

```
PN EP1016710-A2.
XX
PD 05-JUL-2000.
XX
PF 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
XX
XX WPI; 2000-414802/36.
DR P-PSDB; B01786.
DR
XX
XX Increased production of L-amino acids by an Escherichia bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX
XX Disclosure; Page 17-18; 29pp; English.
XX
XX The present sequence is the yahn gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in lysine, glutamic acid and proline is
CC achieved if multiple copies of the gene are transfected into a bacterium.
CC The bacterium used is E. coli.
XX
XX Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;
SQ

alignment_scores:
Quality: 261.50 Length: 206
Ratio: 1.937 Gaps: 3
Percent Similarity: 65.534 Percent Identity: 28.155

alignment_block:
US-09-466-935-4 x A52688 ..
Align seg 1/1 to: A52688 from: 1 to: 672

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProgl 20
58 GTTACCTGACCGTAGGACTG...TTCGTGATTACTTTTTTATCCGG 104

20 yProAspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
105 AGCCAATCTCTTGTGTAGTACAAACAGCGCTGGCTTCCGTCGACGCG 154

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
155 CAGGGGTGCTGACCGGCTGGCGGTGGCGGTGGCGGTGGCGGTGGCGGT 204

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLe 70
205 GGGTGGGTTTGTGGTCTTGCACACGCTAATTACGAGTGTGAGGAGAT 254

70 uHisThrLeuIleMetValGlyGlyLeuTyrLeuCysTrpMetGlyT 87
255 TTTTTCGCTTATCAGAATCGTCGCGGCGCTTATCTCTTATGTTGGGT 304

87 yrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaProAla 103
305 GGTGACGATGCGC.....CGCCAGTCAACA 330

104 ProGlnValGluLeuAlaLysSerGlyArgSer..... 114
331 CCGCAATGAGCAGCTACACACCGATTAGCGCCCTGGTATGCTTT 380

115 .PheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIle 131
```



```

587 CTGACATCGATCATTTAAACCTGTCGCCAGGCTCTGGTGCATCAACAC 636
26 lserGlnThrAlaValSerArgSerArgLysGluAlaMetMetGlyValL 43
637 TATGACCACTCGCTCAACACGGTTATCCGGCGGT ..... 673
43 euGlyIleThrCysGlyVal.....Met 50
674 ..GGCGTCTATGCTGGCTCTAGACCGGACTGGCGATTCATATGCTGT 721
51 ValTPrAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIle.....G1 65
722 GGTTCGGCTGGGTGGGACGCTATTTCCCGCTCAGTGATTGCGTTGA 771
65 uLysMetAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrIL 82
772 AGTGTGAAGTGG.....GCAGGCGCGCTTACT 800
82 euCysTrpMetGlyTyrGlnMetLeuArg.....GlyAlaLeuLysLys 96
801 TGATTGCTGGGAATCCAGCAGTGGCGCGCTGTGCAATTGACCTT 850
97 GluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSerGlyAr 113
851 AAATCGCTGGGCTCTACTCAATCCGCTCGACATTG..... 886
113 gSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIle 130
887 ....TTCCAGCGCGCAGTGTGTGTAATCTCACCAATCCCAAAAGTATG 932
130 leTyrPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThr 146
933 TGTTTCTGGCGGCGCTATTTCGCAATTCATCATGCGCGCACGCCCAA 982
147 ThrAlaArgTrpGlyIlePheAlaLeuIleIleValGluThrLeuAlaTr 163
983 CTGATGCGAGTATCGTCTCGGCTCACC..... 1012
163 pPheThrValValAlaSerLeuPheAlaLeuProGlnMetArgGlyT 180
1013 ....ACTATTGTGGTGNATATTATGTGATGATCGGTACGCCACCTTG 1058
180 yrGlnArgLeuAlaLysTrpIleAspGly..... 189
1059 CTCACCGGATTGCTCTATGGATTAAAGCACCACAGCAGATGAAGCGCTG 1108
190 .....PheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIle 204
1109 AATAAGATTTCGGCTCGCTGTTTATGCTGGTGGGCGGCTGTATGATC 1158
204 eSerArg 206
1159 GCGGAGG 1165

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T06767
seq_documentation block:
XX T06767 standard; DNA; 720 BP.
XX AC T06767;
XX DT
XX DE Mel-linked mlga gene.
XX KW Marine mela; selectable marker; oyster larva settlement;
XX KW pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
XX KW mlga gene; ss.
XX OS Shewanella colwelliana strain lSTDYF.
XX Key Location/Qualifiers

```

```

FT RBS 154..159
FT /*tag= a
XX US5474933-A.
XX PD 12-DEC-1995.
XX PF 21-MAR-1990; 90US-0496804.
XX PR 08-NOV-1993; 93US-0148945.
XX PR 21-MAR-1990; 90US-0496804.
XX PR 10-NOV-1992; 92US-0974837.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX Fuqua WC, Weiner RM;
XX WPI; 1996-039515/04.
XX P-PSDB; R87527.
XX Novel gene encoding marine mela from Shewanella - useful as
PT selectable marker in genetic engineering and for inducing larval
PT oyster settlement
XX Example 7; Fig 14; 47pp; English.
XX The 5' end of an open reading frame (T06767) was identified
CC directly downstream of the Shewanella colwelliana mela gene
CC (T06765), and was designated mlga (mel-linked gene). The
CC role of the encoded protein (R87527) was unclear as deletion
CC subcloning in E. coli demonstrated that only mela was required
CC for melanogenesis.
XX Sequence 720 BP; 180 A; 165 C; 165 G; 210 T; 0 other;

alignment_scores:
Quality: 111.50 Length: 203
Ratio: 0.987 Gaps: 7
Percent Similarity: 55.665 Percent Identity: 25.123

alignment_block:
US-09-466-935-4 x T06767 ..
Align seg 1/1 to: T06767 from: 1 to: 720

21 ProAspPhePheValSerGlnThr..... 29
128 CCGACCTTTTTCGCTCTCATTTACACCGAGGTATGTGTATGACACTGGC 177
30 .....AlaValSerArgSerArgLysGluAlaMetMetGlyValL 43
178 CATGACTCTCGGTATGATGCTGTCGCCGCAACCTTATGGATGATGG 227
43 euGlyIleThrCysGlyValMetValTrpAlaGlyIleAlaLeuLeuGly 59
228 TTGCTGAGCTAGCAGCGTTCGCTCGGCGGATTCGCGCGTAAATGGGT 277
60 LeuHisLeuIleGluLysMetAlaTrpLeuHisThrLeuIleMetVa 76
278 GTCGCCAGTATGATGCTCAACTATCCACAACCTCTTCGATATTTAAATG 327
76 lGlyGlyGlyLeuTyrLeuCysTrpMetGlyTyrGlnMetLeuArgGlyA 93
328 GGTGCGTGGGCTCTATCTTGGTTACATCGGCATTAGCATGTGCGGGCCA 377
93 lateuLysLysGluAlaValSerAlaProAlaProGlnValGluLeuAla 109
378 AAGGAAATGGCCAACTTGCATAATACCTCCAGTCAGATC..... 418
110 LysSerGlyArgSerPheLeu...LysGlyLeuLeuThrAsnLeuAlaAs 125
419 ...AGTAATCGAGCGCTAATAACTCAAGGCTTTGTCCACCGCAATTGCTAA 465

```


ID T96816 standard; DNA; 2374 BP.
 AC T96816;
 XX
 XX 12-MAR-1998 (first entry)
 XX
 XX DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.
 XX
 XX LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
 KW Microbial production; amino acid; animal feed additive; ds.
 XX
 XX *Corynebacterium glutamicum*.
 OS
 FH Key Location/Qualifiers
 FT CDS complement (82..954)
 FT /*tag= a
 FT /label= LysG
 FT 1016..1726
 FT /*tag= b
 FT /label= LysE
 FT complement (1723..2373)
 FT /*tag= c
 FT /label= orf3
 XX
 XX DE19548222-AL.
 XX
 XX 26-JUN-1997.
 XX
 XX 22-DEC-1995; 95DE-1048222.
 XX
 XX 22-DEC-1995; 95DE-1048222.
 XX
 XX (KERJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 XX Eggeling L, Sahm H, Vrije M;
 PI
 DR WPI: 1997-333867/31.
 DR P-PSDB; W3714-16.
 XX
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT *Corynebacterium*
 XX
 PS Claim 23 and 26; Page -; 16pp; German.
 XX
 CC This DNA, isolated from *Corynebacterium glutamicum*, contains the LysG,
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine,
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.
 CC NB This sequence has been created from the information given in table 2
 CC of the specification.
 XX
 XX Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

alignment_scores:
 Quality: 108.00 Length: 263
 Ratio: 0.900 Gaps: 13
 Percent Similarity: 45.627 Percent Identity: 23.574

alignment_block:
 US-09-466-935-4 x T96816 ..

Align seg 1/1 to: T96816 from: 1 to: 2374

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17

```

1022 ATCATGGAAATCTTCATTACAGGTCTGCTTTTGGGGCCAGTCTTTTACT 1071
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgs 34
1072 GTCCATCGACCGCAGAGATGTACTGGTGATTAAACAAGGAATTAAGCGC. 1120
34 exArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
1121 .....GAGGACTCATTCGGGTCTCTCGGTGTGTAAATTTCTGAC 1162
51 ValTrpAlaGlyIleAla...LeuLeuGlyLeuHisLeu..... 62
1163 GTCTTTTGTTCATCGCGGCACCTTGGCGGTGATCTTTTGTCCAATGC 1212
63 .....IleIleGluLysMetAlaTrpLeuHisThrLeuIleMetV 76
1213 CGCGCGCATCGTGCATATATATGCGCTGG..... 1243
76 alGlyGlyLeuTyrLeuCysTrpMetGlyTyrGlnMetLeuArgGly 92
1244 ..GGTGGCATCGCTTACCTGTATGGTTTGGCGTCATGGCAGCAAGAGAC 1291
93 AlaLeuLysLysGluAlaValSerAlaProAlaProGlnVal..... 106
1292 GCCATGACAAACAAG.....GTGGAAGCGCCACAGATCTTGAAGA 1332
107 .....GluLeuA 109
1333 AACAGAACCAACGTCGCCGATGACACGCTTTGGGCGGTTCGCGGTGG 1382
109 laLysSerGlyArgSer..... 114
1383 CCACGTGACACGCGCAACCGGGTGGAGGTGAGCGTCGATAAGCAG 1432
115 .....PheLeuLysGlyLeuLeuThrAsnLeuAlaAs 125
1433 CGGGTTGGGTAACGCCCATGTGTATGGCAATCGTGTGACCTGGTTGAA 1482
125 nProLys.....AlaIleIleTyrPheGlySerValPheSerL 138
1483 CCGAATGCGTATTTGGACGGGTTGTGTATTCGCGGC..... 1522
138 euPheValGlyAspAsnValGlyThrThrAlaArgTrpGlyIlePheAla 154
1523 .....GTGCGCGCAATACGCGCACCGGACGCGTGTGATTTTCGCGCT 1567
155 LeuIleValGluThrLeuAlaTrpPheThrValValAlaSerLeuPh 171
1568 GGCGCGTTCGCGCAAGCCTGATCTGGTCCGCTGGTGG.....TT 1610
171 e.....AlaL 173
1611 TCGGCGCAGCAGCATTTGTCACGCCGCTGTCCAGCCCAAGGTGTGCGCG 1660
173 euProGlnMetArgArgGlyTyr.....GlnArgLeuAlaLysTrpIle 187
1661 TGTATCACTGCTGCTGGCAGGTGTGTGATGACCGCATGGCCATCAACT 1710
188 AspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
1711 GAT...GTTGATGGGTAGTTTTCGCGGTTTGAATC 1746

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52691

seq_documentation_block:
 ID A52691 standard; DNA; 636 BP.

XX AC A52691;
 XX
 XX DT 03-JAN-2001 (first entry)
 XX
 XX DE Escherichia coli yggA gene.


```

457 CTGGTGGAAAGGATCTCTCTCTCTCTCAACAATATGCTCAGTGGG 408
63 ...IlelleGluLysMetAlaTrpLeuHisThrLeuIleMetValGlyG1 78
407 GATGACGCTTTTATGATACATATGATTCACACAGCTGGACATTCGGGG 358
78 YGlyLeuTyLeuCysTrpMetGlyTyGlnMetLeuArgGlyAlaLeuL 95
357 G.....TGGGTT.....AGGATGGTGG..... 339
95 YsLysGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
338 .....CCCATGGCAATGTTTCATCATATGTTGTCG 309
112 GlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAl 128
308 TTCAGACCG.....AAGGCTCTGCCGAAGAC 283
128 allelleTyPheGlySerValPheSerLeuPheValGlyAspAsnValG 145
282 GGTCAAG.....CAGCTTTTCCGCTGTCTGTTGTTTGCAGTAG 242
145 ly Thr.....ThrAlaArgTrpGlyIlePheAlaLeuI 156
241 GCCACACACACATGATGCCGACCAAGAGGGCGATGCAGATGCCGGT 192
156 elleValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaL 173
191 TATGGTCAGCACTCTCTTGTGACAGCTCCCTCCGCCCTCAATAATCAA 142
173 eupProGlnMetArgGlyTyGlnArgLeuAlaLysTrpIleAspGly 189
141 TCCCA.....AGATGCTTTTCTTGTTGGTGGACT 113
190 PheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeu 202
112 TTCTATGGGCACA.....TTCTCAGTACATCTT 86

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:C74616
seq_documentation_block:
ID C74616 standard; cDNA; 401 BP.
AC C74616;
DE 08-FEB-2001 (first entry)
DE Human ORFX ORF171 polynucleotide sequence SEQ ID NO:341.
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW immunostimulant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW hypotensive; thrombolytic; coagulant; vasotropic; antidiabetic;
KW antitumor; antibacterial; immunosuppressive; antinflammatory;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
KW thrombosis; contraceptive; ss.
OS Homo sapiens.
PN WO200058473-A2.
PD 05-OCT-2000.
PF 31-MAR-2000; 2000WO-US08621.
XX
```

```

PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 03-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI: 2000-602362/57.
DR P-PSDB; B40407.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 607-608; 5507pp; English.
XX
XX C74446 to C77606 encode the proteins given in B40237 to B43397, which
XX represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
XX Sequence 401 BP; 57 A; 121 C; 124 G; 98 T; 1 other;
XX
```

```

alignment_scores:
Quality: 94.50 Length: 136
Ratio: 1.277 Gaps: 6
Percent Similarity: 54.412 Percent Identity: 25.735

alignment_block:
US-09-466-935-4 x C74616 ..
Align seg 1/1 to: C74616 from: 1 to: 401

81 TyrLeuCysTrpMetGlyTyGlnMetLeuArgGlyAlaLeuLysLysG1 97
|||||.....:|||||
13 TACCTGCTGTACCTGGCTATGCCACCTGGCGTAC.....CGCTGGC 56
97 uAlaValSerAlaProAlaProGlnValGluLeuAlaLysSerGlyArgS 114
:::.....:|||||
57 CTTTGCATGAACGACACGCCGACAGTTGCCACGCCGCGCAGC.....C 100
114 erPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIle 130
:::|||||.....:|||||
101 TGATCCTGCGCTCTCTGCTGACATCTTAAACCCCAAGCTGACAATT 150
131 TyrPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThrTh 147
:::|||||.....:|||||
151 TTCTTCTGCGCTTCTGCTGCTCAATTCGTA.....ACGCCAGCGCGCAC 194
147 rAla.....ArgTrpGlyIlePheAlaLeuIleI 157
|||||.....:|||||
195 CGCGCGCGCTTCAGATGCTGCTACTGAGCGCGCTGTTTCATGGCGATGA 244
157 leValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaLeu 173
```

```

245 CGCTTCAGTGTGTTGCTGATGGCTGTGGCAATGTGTTT..... 288
174 ProGlnMetArgArgGlyTyr.....GlnArgLeuAlaLysPrp1l 187
289 .....CGTCGTGAGTGGTCGAGTCGCCAGGTGTCAGAACTGGCT 329
187 eAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIleI 204
330 CGGAGCGAGTTTGGCCAGGGCTTTGGCCGGGCTGGGGTTGAACCTGGCGT 379
204 leSerArg 206
380 TTGCGCAG 387
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:233622

seq_documentation_block:

ID 233622 standard; cDNA; 1597 BP.

XX Z33622;

DT 08-DEC-1999 (first entry)

DE Human breast tumour-associated EST 12.

XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;

KW treatment; tumour; cytostatic; medicament; ss.

XX Homo sapiens.

XX DE19813839-Al.

XX 23-SEP-1999.

PF 20-MAR-1998; 98DE-1013839.

PR 20-MAR-1998; 98DE-1013839.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;

DR WPI; 1999-528981/45.

XX Human nucleic acid sequences and protein products from tumor breast

PS tissue, useful for breast cancer therapy -

XX Claim 3; 93; 188pp; German.

CC This invention describes novel human nucleic acid sequences from tumor
CC breast tissue which have cytotatic activity. The nucleic acid sequences
CC can be used to produce and isolate full-length gene sequences. They can
CC be used to express proteins, which can be used as tools to find an
CC activity against breast cancer. The sequences can be used in sense or
CC antisense form. They are especially useful for medicaments for gene
CC therapy to treat breast cancer. 233611-248617 represents expressed
CC sequence tags described in the method of the invention.

XX Sequence 1597 BP; 476 A; 229 C; 337 G; 555 T; 0 other;

alignment_scores:

Quality: 90.50 Length: 180
Ratio: 1.077 Gaps: 11
Percent Similarity: 46.567 Percent Identity: 27.778

alignment_block:

US-09-466-935-4 x Z33622 ..

Align seg 1/1 to: Z33622 from: 1 to: 1597

53 AlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMetAlaTr 69

```

29 GCGGCGTCGCCACTCTG..... 111111
69 pLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCysTyrMetG 86
52 G.....ATGGTGATATGGGTGATGCCCGCACAAATTCACGGATG 92
86 lyTyrGlnMetLeuArgGlyAlaLeuLysGluAlaValSerAla... 101
93 GCCTAGCAAT.....GGTGCCTCTTTTACTGAGGCTTATCAAGTGT 136
102 .....ProAlaProGln..... 105
137 TTAAGTACTTCTGTGCTGTCTGTGTCATGAGTTGCCCTCATGAATTAGG 186
106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuL 120
187 TGACTTGTCTTCTACTAAGGCTGGCATGACCGTTAAGCAGGCTGCC 236
120 euThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValPhe 136
237 TTTATAAT...GCATTGTCAGCCATGCTGCCGTATCTTGAATGGCAACA 283
137 SerLeuPheValGlyAspAsnValGlyThrThrAlaArgTyrPglyIlePh 153
284 GGAATTTTCATTGCTCATTTATGCTGAAATGTTCTATGTGG...ATATT 330
153 eAlaLeuIleIleValGluThrLeuAlaTrpPheThrValValAlaSerL 170
331 TGCACCTT.....ACTGCTGCTTATTCATGTATGTGCTCTGG 368
170 euPheAlaLeuProGlnMet..... 176
369 TTGATATGTGTACCTGAAATGCTGCACAATGCTAGTGACCATGGATGT 418
177 ...ArgArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPheAlaGl 192
419 AGCCGCTGGGGTATTTCTTTTACAGAAAT.....GCTGG 453
192 yAlaLeuPheAlaGlyPheGlyIleHisLeuIleIleSer 205
454 GATGCTTTTG...GTTTGGAAATATGTTACTTATTTC 490
```

```
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/847,743B
; FILING DATE: 19920306
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705256
; FILING DATE: 24-May-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/765212
; FILING DATE: 25-Sep-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/790801
; FILING DATE: 08-No. 5367060-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Hensley, Max D.
; REGISTRATION NUMBER: 27,043
; REFERENCE/DOCKET NUMBER: 712P3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/266-1489
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-847-743B-7

alignment_scores:
  Quality: 90.00      Length: 199
  Ratio: 0.938       Gaps: 12
  Percent Similarity: 48.241  Percent Identity: 26.131

alignment_block:
US-09-466-935-4 x US-07-847-743B-7/rev ..

Align seq 1/1 to reverse of: US-07-847-743B-7 from: 1 to: 2199

20 GlyProAspPhePheValSer...GlnThrAlaValSerArgSerAr 35
1202 GGAAGAGGATGCTTTTCAGTGTCTCCGTCCTCCACGCTGGCTAGAGTCT 1153
35 glysGluAlaMetMetGlyValLeuGlyThrCysGlyValMetValT 52
1152 GGGTGCACACT...ACTGGAGTGCATGGGC.....TGTGAAGTATAGTA 1112
52 rPalaglytleAlaLeuLeuGlyLeuHisLeu..... 62
1111 CTGGTGGAAAGGAGTGTCTCTGCTCTCTCAACAATATGCTCACTGGA 1062
63 ...IleIleGluLysMetAlaTrpLeuHisThrLeuIleMetValGlyG 78
1061 GATGACGTTTTAGATACGATGATTGATTCACCGAGCTGGACATCTCGGG 1012
78 yGlyLeuTyrrLeuCysTrpMetGlyTyrrGlnMetLeuArgGlyAlaLe 95
1011 G.....TGGGTT.....AGGATGGTGAGG..... 993
95 yLysGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
992 .....CCCATGGCAATGTTTCATCATATGTTTTCG 963
112 GlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAl 128
962 TTCAGACCG.....AAGGCTCTGCGGAAGAC 937
128 alletIleTyrrPheGlySerValPheSerLeuPheValGlyAspAsnVal 145
936 GGTGATG.....CAGCTTTTCCGCTGTTCTTCTTGGTTTTCAGTAG 896
```

```
145 ly.Thr.....ThrAlaArgTrpGlyIlePheAlaLeuIl 156
|| ||| ||||| ||||| ||||| ||||| ||||| |||||
895 GCCACCACACACATGATGCGGACACACAGAGGGCGATGCGCGGT 846
156 eileValgluThrLeuAlaTrpPheThrVal1ValAlaSerLeuPheAlaL 173
||||| ||||| ||||| ||||| ||||| ||||| |||||
845 TATGGTCAGCAGCTCTCTTCTGGTACAGCTCTCCGCGCTCCATAAATTC 796
173 euProGlnMetArg..... 177
795 TCCACAGATGCTTGTAGAAAGTGGCCATTACGTAGTTTGGCAGCGATCA 746
178 .....ArgGlyTyrrGlnArgLeuAlaLysTrpIleAspGlyPhe 190
||||| ||||| ||||| ||||| ||||| ||||| |||||
745 CCAGTAACTCATTTGGGCACCTTGACACAGATATCTCGAGGGGTTT 701

seq_name: /cgn2.6/ptodata/2/1na/5A_COMB.seq:US-08-096-277-7

seq_documentation_block:
; Sequence 7, Application US/08096277
; Patent No. 5578482
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckeet
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/096,277
; FILING DATE: 26-JUL-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/875,788
; FILING DATE: 29-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/640,497
; FILING DATE: 14-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,988
; FILING DATE: 24-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,114
; FILING DATE: 22-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/528,438
; FILING DATE: 25-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale H
; REGISTRATION NUMBER: 19,090
; REFERENCE/DOCKET NUMBER: 02899,43360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
```


seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-550-815-7

seq_documentation_block:

; Sequence 7, Application US/08550815

; Patent No. 5869618

; GENERAL INFORMATION:

; APPLICANT: Lippman, Marc E

; APPLICANT: Lupu, Ruth

; TITLE OF INVENTION: Ligand Growth Factors that Bind to the

; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response

; NUMBER OF SEQUENCES: 18

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner, Birch, McKie & Beckett

; STREET: 1001 G Street, N.W.

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.

; ZIP: 20001

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/550,815

; FILING DATE: 31-OCT-1995

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/096,277

; FILING DATE: 26-JUL-1993

; APPLICATION NUMBER: US 07/875,788

; FILING DATE: 29-APR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/640,497

; FILING DATE: 14-JAN-1991

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/917,988

; FILING DATE: 24-JUL-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,114

; FILING DATE: 22-APR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/528,438

; FILING DATE: 23-MAY-1990

; ATTORNEY/AGENT INFORMATION:

; NAME: Hoscheit, Dale H

; REGISTRATION NUMBER: 19,090

; REFERENCE/DOCKET NUMBER: 02899,43360

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-508-9100

; TELEFAX: 202-508-9299

; TELEX: 197430 BBMB UT

; INFORMATION FOR SEQ ID NO: 7:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2199 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: both

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA (genomic)

; US-08-550-815-7

alignment_scores:

Quality: 90.00 Length: 199

Ratio: 0.938 Gaps: 12

Percent Similarity: 48.241 Percent Identity: 26.131

alignment_block:

US-09-466-935-4 x US-08-550-815-7/rev ..

Align seg 1/1 to reverse of: US-08-550-815-7 from: 1 to: 2199

```

20 GlyProAspPhePheValSer...GlnThrAlaValSerArgSerAr 35
1202 GGAAGAGTGTTCAGTGTCCGTTGCTCCAGCTGGCTAGAGTCT 1153
35 gLySGluAlaMetMetGlyValLeuGlyLleThrCysGlyValMetValT 52
1152 GGGTGACAGT...AGTGGAGTGATGGGC.....TGTGGAAGTATAGTGA 1112
52 rPalagLyLeAlaLeuLeuGlyLeuHisLeu..... 62
1111 CTGGTGGAAAGGATGCTCTGCTTCTCTCAACAATATGCTCACTGGA 1062
63 ...llelleGluLysMetAlaTrpLeuHisThrLeuLeuMetValGlyGl 78
1061 GATGAGCTTTTAGATAGTATTCATTCACCAAGCTGGACATTCCTCGGGG 1012
78 yGlyLeuTyrLeuCysTrpMetGlyTyrGlnMetLeuArgGlyAlaLeuL 95
1011 G.....TGGGT.....AGGATGGTCAGG..... 993
95 ySLySGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
992 .....CCATTGGCAATGTTTCATCATATGTTTCG 963
112 GlyArgSerPheLeuLysGlyLeuThrAsnLeuAlaAsnProLysAl 128
962 TTCAGACCG.....AAGGCTCTGCCAAGAC 937
128 atlelleTyrPheGlySerValPheSerLeuPheValGlyAspAsnValG 145
936 GGTCAAG.....CAGCTTTTCCGCTGTTCTTCGTTTTCAGTAG 896
145 ly.Thr.....ThrAlaArgTrpGlyllePheAlaLeuL 156
895 GCCACCACACATGATGCCGCCAAGGAGGGCGATGCAGATGCCGT 846
156 eileValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaL 173
845 TATGGTCAGCACTCTCTCTGTTAGTACAGCTCCTCGGCTCCATAAATCAA 796
173 euProGlnMetArg..... 177
795 TCCCAAGATGCTTGTAGAGCTGCCATTACGTAGTTTGGCAGCGATCA 746
178 .....ArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPhe 190
745 CCAGTAAACTCATTTGGGCACATTGCACAAGTATCTCGAGGGGTTT 701
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-703-089-7
seq_documentation_block:
; Sequence 7, Application US/08703089
; Patent No. 6040290
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20001
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US-08-703,089
; FILING DATE: 31-OCT-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,277
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: US 07/875,788
; FILING DATE: 29-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/640,497
; FILING DATE: 14-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,988
; FILING DATE: 24-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,114
; FILING DATE: 22-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/528,438
; FILING DATE: 23-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale H
; REGISTRATION NUMBER: 19,090
; REFERENCE/DOCKET NUMBER: 02899,43360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-550-815-7

```



```

125 snProlysAlaIleIleIleTyrPheGlySerValPheSerLeuPheValGly 141
||||| :|||: :|||: ||| :|||
4756 ACCACAC.....GTCTACCTGCACACCGTCGTCGTAGCGCGCTG 4713
||||| :|||: :|||: ||| :|||
142 AspAsnValGlyThrAlaArgTrpIlePheAlaLeu...IleI 157
||||| :|||: :|||: ||| :|||
4712 GCCACGACGACGACGACGCTGG...CTGTGCGCCCTCGCGCGGT 4666
||||| :|||: :|||: ||| :|||
157 eValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaLeuP 174
||||| :|||: :|||: ||| :|||
4665 CACAGCCAGTCGGTATGGTTCGCCACCCCTCGG.....TTCGGAGCG 4622
||||| :|||: :|||: ||| :|||
174 roGlnMetArgArgGlyTyrGlnArgLeuAlaLysTrp.....IleAsp 188
||||| :|||: :|||: ||| :|||
4621 GCCGGTTCGGCGGCTGTTCCACCAACCCGCGCTCGTGAGAAATCTCGAC 4572
||||| :|||: :|||: ||| :|||
189 GlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIleIle 205
||||| :|||: :|||: ||| :|||
4571 GGCTGATCGCGTCATGATGTTGCGCTGGGAATCTCGCTGACCGTGAC 4522
205 r 205
4521 c 4521
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-311-023-1

```

```

seq_documentation_block:
; Sequence 1, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1310 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1293
US-08-311-023-1

```

```

alignment_scores:
Quality: 86.50 Length: 179
Ratio: 1.042 Gaps: 11
Percent Similarity: 46.369 Percent Identity: 27.374

alignment_block:
US-09-466-935-4 x US-08-311-023-1 ..

Align seg 1/1 to: US-08-311-023-1 from: 1 to: 1310
53 AlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGlyLysMetAlaTr 69
|||||:|||||
787 GCCGGCGTGCACACTTG.....GCCGTG 809
|||||:|||||
69 pLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuCysTrpMetG 86
|||||:|||||
810 G.....ATGGTGATTAATGGGTGATGGCTGCACAAATTCAGCGCATG 850
|||||:|||||
86 lyTyGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSerAla... 101
|| :|||:|||||
851 GCCTAGCAATT.....GGTCTGCTTTTACTGAAGCTTATCAAGTGT 894
|||||:|||||
102 .....ProAlaProGln..... 105
|||||:|||||
895 TTAAGTACTTCTGTGCTGTCTGTGCATGAGTTGCCTCATGAATTAGG 944
|||||:|||||
106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuL 120
|||||:|||||
945 TGACITTTGCTGTTCTACTAAAGGTGGCATGACCGTTAAGCAGCGTGTCC 994
|||||:|||||
120 euThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValPhe 136
|| ||| ||| :|||:|||||
995 TTATAAT...GCATTGTACCCCATGCTGGCGTATCTTGAATGGCAACA 1041
|||||:|||||
137 SerLeuPheValGlyAspAsnValGlyThrAlaArgTrpGlyIlePh 153
|||||:|||||
1042 GGAATTTTCATTGGTCATTATGCTGAAATGTTTCTATGTGG...ATATT 1088
|||||:|||||
153 eAlaLeuIleIleValGluThrLeuAlaTrpPheThrValValAlaSerL 170
|||||:|||||
1089 TGCACIT.....ACTGCTGGCTTATTCATGATGTTGCTCTCG 1126
|||||:|||||
170 euPheAlaLeuProGlnMet..... 176
|| :|||:|||||
1127 TTGATATGGTACTGAAATGCTGCACAAATGATGCTAGTGACCATGGATGT 1176
|||||:|||||
177 ...ArgArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPheAlaG 192
|||||:|||||
1177 AGCCGCTGGGGGTATTTCTTTTACAGAAT.....GCTGG 1211
|||||:|||||
192 yAlaLeuPheAlaGlyPheGlyIleHisLeuIleIle 204
|||||:|||||
1212 GATGCTTTTG...GGTTTGAATATGTTACTTATT 1245
|||||:|||||

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-311-023-3
seq_documentation_block:
; Sequence 3, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA

```

```
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32,925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-311-023-3

alignment_scores:
    Quality: 86.50      Length: 179
    Ratio: 1.042      Gaps: 11
    Percent Similarity: 46.369      Percent Identity: 27.374

alignment_block:
US-09-466-935-4 x US-08-311-023-3 ..

Align seg 1/1 to: US-08-311-023-3 from: 1 to: 2404

53 AlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMetAlaTrp 69
|||||:|||||
787 GCCGGCTGCGCAGCTTG.....GCCTG 809

69 pLeuHisThrLeuIleMetValGlyGlyLeuTyTrpMetG 86
|:|||||:|||||
810 G.....ATGGTATAGGTGATGCCGTCACAAATTCACCGATG 850

86 lyTrpGlnMetLeuArgGlyAlaLeuLysGluAlaValSerAla... 101
||:|||||:|||||
851 GCCTAGCAAT.....GGTCTGCTTTTACTGAAGGCTTATCAAGTGT 894

102 .....ProAlaProGln..... 105
|||||:

895 TTAAGTACTTCTGTGTCTCTCTCATGATTCGCTCATGAATTAGG 944
|||||:

106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuL 120
||:|||||:|||||
945 TGACTTTGCTGTTCTACTAAGGCTGGCATGCCGTTAAGCAGGCTGCC 994

120 euThrAsnLeuAlaAsnProLysAlaIleIleTyPheGlySerValPhe 136
||:|||||:|||||
995 TTTAATAAT...GCATTGTCAGCATGCTGCCGTATCTTGAATGCAACA 1041

137 SerLeuPheValGlyAspAsnValGlyThrAlaArgTrpGlyIlePh 153
|||||:|||||
1042 GGAATTTTCATGTCATTATGCTGAAATGTTCTATCTGG...ATATT 1088

153 eAlaIleIleValGluThrLeuAlaTrpPheThrValValAlaSerL 170
|||||:|||||
1089 TGCACCT.....ACTGCTGGCTATTTCATGATGTGTCGTGG 1126

170 euPheAlaLeuProGlnMet..... 176
||:|||||:|||||
1127 TTGATAGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGT 1176

177 ...ArgArgGlyTyTrpGlnArgLeuAlaLysTrpIleAspGlyPheAlaG 192
||| ||||| ||| :|||
1177 AGCGCGCTGGGGTATTCTTTTACAGAAT.....GCTGG 1211

192 yAlaLeuPheAlaGlyPheGlyIleHisLeuIleIle 204
| |||||: ||||| |||||
1212 GATGCTTTG...GGTTTGGAAATTATGTTACTATT 1245

seq_name: /cgn2_5/ptodata/2/lna/5A_COMB.seq:US-08-395-246C-1

seq_documentation_block:
; Sequence 1, Application US/08395246C
; Patent No. 5773214
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; TITLE OF INVENTION: ASPERGILLUS FLAVUS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/395,246C
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35784
; REFERENCE/DOCKET NUMBER: x9683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; TELEFAX: 317-277-1917
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3924 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3924
; US-08-395-246C-1

alignment_scores:
    Quality: 85.50      Length: 173
    Ratio: 1.082      Gaps: 8
    Percent Similarity: 45.665      Percent Identity: 23.699

alignment_block:
US-09-466-935-4 x US-08-395-246C-1 ..

Align seg 1/1 to: US-08-395-246C-1 from: 1 to: 3924

3 MetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerPr 19
|||||:|||||
2347 CTCATGTTCTCATGATGGCATGTGTCGCTGTGTGTATAGTGCGCA 2396

19 oGlyProAspPhePhePheValSerGlnThrAlaValSerArgL 36
||| ||| :|||
2397 AGGACCTTGTGTGCATATAGTTCGAAAAGATGCTCTACCGGCCGCA 2446
```

```
36 yscLuAla..... 38
2447 GTCAAGCATTCGGGTGATCTACTGACACAGGATATCTTTCTTCGATCAA 2496
38 ..... 38
2497 CAGGAAACACGACAGCGGCATTAAACAGCCACTCTGAGCGCGGGACCAA 2546
39 ....MetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaG 54
2547 AGAACTAACTGACATAGTGGCGTCACATTGGGACTATCTTGATTGTCT 2596
54 ly.....IleAlaLeuLeuGlyLeuHisLeuIleGluLysMet 57
2597 CTTGTCATCTGTGGCTTCCTTGGGGTAGCCCTTGTAATA..... 2637
68 AlaTrpLeuHisThrLeuIleMetValGlyGly.....GlyLeuTrpLe 82
2638 GGCTGGAAGTTGGCACTGGTGTGATCTCCCGGTTTCCAGCCCTGCTGAT 2687
82 uCys.....TrpMetGlyTyTrpGlnMetLeuArgGlyAlaL 94
2688 GTGTGATTTGTTCGCTTGGTGTGGAGCGGTTCCACGAGGCGCA 2737
94 euLys.....LysGluAlaValSerAla.ProAlaTrp 104
2738 AGNAGCATATCAAGATCACCTAGTTCTGTTGTGAGCGACCTCTGCC 2787
104 oGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuT 121
2788 ATCCGTA.....CGGTGCTTCTTTGACCATGAGGAC 2819
121 hrAsnLeuAlaAsnPro 126
2820 GGAGGCTTGTCAATCCT 2836
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-335-409-1
seq_documentation_block:
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
alignment_scores:
Quality: 82.00 Length: 209
Ratio: 0.845 Gaps: 9
Percent Similarity: 46.411 Percent Identity: 22.010
alignment_block:
US-09-466-935-4 x US-09-335-409-1 ..
Align seq 1/1 to: US-09-335-409-1 from: 1 to: 68750
```

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuWe 17

```
3439 GTCCCTGCTCTGCTGTCACCCCTCGCGCTCATCTCTGTCGACGCGCGC... 3486
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
3487 .....GCCTCCGCGGAGCTCGCGGGCGGC 3511
34 erArgLysGluAlaMetMetGlyValLeu.....GlyIleThrCysGly 48
3512 TCGCCACCGCCGAGCTGCTCGGGAGCTCTCGCGGCGCTGCTGCTGGC 3561
49 ValMetValTrpAlaGlyIleAlaLeuLeuGlyLeuHis..... 61
3562 CCCTCCGCTC...GTGCGCGGCTCGCGCGGGTTCCATCGAGCCCTCTT 3608
62 .....LeuIleIleGluLysMetAlaTrpLeuHisT 72
3609 CAGAGAGCGCGGCTCGGGTCTGCTCTCGGGCATCTCCTGG..... 3651
72 hrLeuIleMetValGlyGlyLeuTrpLeuCysTrpMetGlyTrpGln 88
3652 .....ATAGCGCGCTCTCTCTGCTGATGCGCGGCATCGAG 3690
89 MetLeuArgGlyAlaLeuLysGlyAlaValSerAlaProAlaProG1 105
3691 CTCGACCTGGGCTCTCTGCGCAAGAGCG..... 3720
105 nValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuThra 122
3721 .....CGCCCGCGGCGCTCTCGCGCTCGGCG 3748
122 snLeuAlaAsnProLysAlaIleIleTyPheGlySerValPheSerLeu 138
3749 CGATCGCGCGCCCTCGCGCG.....GGCGCGCTCTCTCGCGC 3789
139 PheValGlyAspAsnValGlyThrThrAlaArgTrpGlyIlePheAlaLe 155
3790 CTCGCTCGATCGCGCCCTCTCGAGC.....GGCCTCTCTCTCGG 3830
155 tIleIleValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheA 172
3831 GATCGTCTCTCGGTGACGCGCGGTGATCGCGAGGTGCTGATCG 3880
172 laLeuProGlnMetArgGlyTyTyr..... 180
3881 AGCGGAGTCTGATCGCGCGAGTATCGCAGGTGACGCTCGCGCGGG 3930
181 .....GlnArgLeuAlaLysTrpIle 187
3931 GTGGTACAGGAGTCTGCTGGTG 3957
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-940-424-10
seq_documentation_block:
; Sequence 10, Application US/08940424A
; Patent No. 6171837
; GENERAL INFORMATION:
; APPLICANT: Blauer, William S.
; APPLICANT: Zott, Roseann P.
; APPLICANT: Gamble, Mary V.
; APPLICANT: Mertz, James R.
; TITLE OF INVENTION: POTENT
; FILE REFERENCE: 0575/54544
; CURRENT APPLICATION NUMBER: US/08/940,424A
; CURRENT FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1229
; TYPE: DNA
; ORGANISM: mouse
US-08-940-424-10
```

```

alignment_scores:
  Quality: 79.00      Length: 209
  Ratio: 0.898      Gaps: 9
  Percent Similarity: 42.105      Percent Identity: 22.488

alignment_block:
US-09-466-935-4 x US-08-940-424-10  ..

Align seg 1/1  to: US-08-940-424-10  from: 1  to: 1229

12  HisIleValAlaLeuMetSerProGlyProAspPhePheValSerG1 28
|||||
93  CACATTGTGTG.....CTGCCAGCTTCCCCAGAGCCT 127
28  nThrAlaValSerArgLysGluAlaMetMetGlyValLeuGlyI 45
:::|||||:::  |||
128  AGCTGCCCTCAGCAGGCGCATCTCATCCCATCATGTGG.....CTGCCCTC 171
45  leThrCysGlyValMetValTPAlaGlyIleAlaLeuLeuGlyLeuHis 61
:::|||||:::|||||:::|||||:::
172  TGCTTCGGGTGCTTGCCTGTGGCGAGTCTGTGCTGTCTCAGACCGG 221
62  LeuIleIleGluLysMetAlaTrpLeuHisThrLeuIleMetValGlyI 78
:::|||||:::|||||:::
222  .....CAGAGCCTGCCGCGCATGTGCTTTCATCTTCATCACTGG 262
78  Y.....GlyLeuTyrLeuCystTrpMetG 86
|  ::|||  |||
263  CTGTGACTCTGGCTTTGGGCGCTTCTGGCACTGCACTTGACCCAGAAG 312
86  lYTrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaPro 102
|||||:::|||||:::|||||
313  GCITCCAGTCCGCGCGCTGCTG.....ACCCC 344
103  AlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLe 119
:::  ::|||  :::  |||
345  TCTGGAGCAGAGACCTGCGAGCATGGCTCTCCCGCCTCCACACAAC 394
119  uLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValP 136
{||  :::|||||:::
395  ACTACTGGATATCATCATCCCCAG..... 419
136  heSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp..... 150
|  |||||  :::|||||
420  .....ATGTCACCAAGTTGCCAAGTGGGTGAAG 449
151  .....GlyIlePheAlaLeuIle..... 156
|  |||:::|||||:::
450  ACACGTCTGGAGAAACTGGACTTTTGGTCTGGTGAATAACGCTGGCGT 499
157  .....IleValGluThrLeuAlaTrpPheThr..... 165
|  |||:::  |||:::
500  AGCTGGTATCATCGGCGCCACACCATGGCTGAACACAGGATGATTCCAGA 549
166  .....ValValAlaSerLeu 170
|  |||:::  ::|||
550  GAGTACTGAGTGTGAACACACTGGGGCCCATCGGTGTACCCCTGGCCCTG 599
171  PheAlaLeuProGlnMetArgGly 179
:::  |||  |||:::
600  CTGCCCTGTACAGCAGGCCAGGGGT 626

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:40:01 ; Search time 26.05 Seconds
(without alignments)
543.452 Million cell updates/sec

Title: US-09-466-935-4
Perfect score: 1054
Sequence: 1 MLMFLTVAMVHIVALMSGP.....IDGFAGALFAGFGIHLIISR 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618	58.6	122	2	H65186
2	368.5	35.0	210	2	F64115
3	313.5	29.7	222	2	F82333
4	292	27.7	210	2	D84016
5	274.5	26.0	204	2	B83279
6	269.5	25.6	212	2	T43921
7	263.5	25.0	213	2	F83444
8	261.5	24.8	223	2	H64759
9	246.5	23.4	204	2	F83306
10	245	23.2	210	2	G83082
11	208.5	19.8	206	2	G82979
12	207	19.6	216	2	F83051
13	201	19.1	207	2	B83703
14	199	18.9	216	2	C55580
15	197.5	18.7	204	2	B82410
16	190	18.0	212	2	F64940
17	186	17.6	208	2	G84086
18	180.5	17.1	210	2	F69975
19	174.5	16.6	241	2	C75329
20	172	16.3	212	2	G82200
21	167.5	15.9	209	2	E82388
22	163	15.5	213	2	C82523
23	159	15.1	203	2	T30317
24	159	15.1	206	2	S76178
25	155.5	14.8	208	2	C82471
26	142.5	13.5	205	2	G82358
27	141.5	13.4	206	2	C82139
28	136	12.9	207	2	D83187
29	123.5	11.7	208	1	B69066

ALIGNMENTS

RESULT 1

H65186
hypothetical 13.3 kD protein in recQ 3' region - Escherichia coli (strain K-12)
N:Alternate names: hypothetical protein ol28
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: H65186; S30713
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.,
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H65186
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-122 <BLAT>
A:Cross-references: GB:AE000458; GB:U00096; NID:92367299; PIDN:AAC76826.1; PID:923673
A:Experimental source: strain K-12. substrain MG1655
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.
A:Reference number: S30560; MUID:92358234
A:Accession: S30713
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'YAVASLLD',2,'LPDATWCTE',12,'RGSFCTCATGRAGEKWAQ',31,'PES',35-122 <DAN>
A:Cross-references: EMBL:M87049
A:Note: this sequence has been corrected
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: ylgJ

Query Match 58.6%; Score 618; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 2.6e-48; Mismatches 0; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 85 MGYOMLRGALKKEASAPAPQVELAKSGRSFLKGLLTNLNLANPKAIYFGSVSFLFVGDNV 144
Db 1 MGYOMLRGALKKEASAPAPQVELAKSGRSFLKGLLTNLNLANPKAIYFGSVSFLFVGDNV 60
Qy 145 GTTARNGIFALIIIVETLAWFTVVASLFALPQMRRGYORLAKWIDGFAGALFAGFGIHLII 204
Db 61 GTTARNGIFALIIIVETLAWFTVVASLFALPQMRRGYORLAKWIDGFAGALFAGFGIHLII 120
Qy 205 SR 206
Db 121 SR 122

RESULT 2

F64115

```
hypothetical protein H11307 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: F64115
R:Fietschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.;
Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.;
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID: 95350630
A:Accession: F64115
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-210 <TIGR>
A:Cross-references: GB:U32810; GB:I42023; NID:g1574760; PIDN:RAC22954.1; PID:g1574766;
C:Genetics:
A:Start codon: GTG
C:Superfamily: hypothetical protein bl798

Query Match 35.0%; Score 368.5; DB 2; Length 210;
Best Local Similarity 38.8%; Pred. No. 8.1e-26;
Matches 80; Conservative 42; Mismatches 73; Indels 11; Gaps 4;

QY 4 LFTVAVMHVIALMSPGDPFFVSGTAVSRSRKEAMGVLCITCGVMWAGIALGLHLI 63
Db 1 MMLNLIIVHFLGMLTPGDPFFVSRMAASNRNTVCGILGITLGIATFWGMSLGLGLAVL 60

QY 64 IEKMAWLHTLIMVGGGLYLCWGYQMLRGA--LKKEAVSAPAPQVELAKS---GRSFLKGL 120
Db 61 FVTIPALHGVIMLLGGSLAVLGLFMARS--KKYKPFESHSDTEFNQQTIKKELKGL 118

QY 121 TNLANPKAIYFGVSFLVGDVNGVTTARWGI---FALLIVETLAWFTVVASLFPALPQMR 177
Db 119 VNLUNAKVIVYESSVMSLVL---VNITEMQOIIIAFAVIVVETFCYFVVISLIFSRNIAK 175

QY 178 RGYQRLAKWIDGFAGALFAGEGIIHLI 203
Db 176 RLYSQYSRYIDNMAGIVFLFGCVLV 201

RESULT 3
conserved hypothetical protein VC0191 [imported] - Vibrio cholerae (group O1 strain N169
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: F82353
R:Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardson, D.; Ermolaeva, M.B.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P.
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID: 20406833
A:Accession: F82353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-222 <HEI>
A:Cross-references: GB:AE004109; GB:AE003852; NID:g9654590; PIDN:AAF3367.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0191
A:Map position: 1

Query Match 29.7%; Score 313.5; DB 2; Length 222;
Best Local Similarity 31.3%; Pred. No. 6.9e-21;
Matches 62; Conservative 52; Mismatches 79; Indels 5; Gaps 2;

QY 2 LMLFTVAVMHVIALMSPGDPFFVSGTAVSRSRKEAMGVLCITCGVMWAGIALGLHLI 61
Db 4 MSILATLAGVHFIALSPGDPVALVQNAQHGKRTGVMTALGLSGILVHLIISLGSIS 63

QY 62 LIIEKMAWLHTLIMVGGGLYLCWGYQMLRGAKEAVS---APAPQVELAKSGRSFLK 117
Db 64 YLVVQOPMFLNLQLAGSSVLLYLGALQSVMAQKNASTPTHSPAPSI-LGNRQRAFTK 122

QY 118 GLLTNLANPKAIYFGVSFLVGDVNGVTTARWGI--FALLIVETLAWFTVVASLFPALPQMR 177
Db 123 GMMNLLNPKALVFFVSLSSLSLIPASMSVSGKVSAAAILVGLSLTWFSCLAWLTTTSAMQ 182

QY 178 RGYQRLAKWIDGFAGALF 195
Db 183 QRMQRITRSVDSICAAYF 200

RESULT 4
D84016
hypothetical protein BH2932 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: D84016
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H.
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID: 20263314
A:Accession: D84016
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BAB06651.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2932
C:Superfamily: hypothetical protein bl798

Query Match 27.7%; Score 292; DB 2; Length 210;
Best Local Similarity 32.4%; Pred. No. 5.4e-19;
Matches 67; Conservative 37; Mismatches 97; Indels 6; Gaps 2;

QY 4 LFTVAVMHVIALMSPGDPFFVSGTAVSRSRKEAMGVLCITCGVMWAGIALGLHLI 63
Db 1 MFVEVFIIVGLLAGMSPGDPFFIVMKNLSLGFARVGLITSLGASALIVHITYTVLGFALF 60

QY 64 IEKMAWLHTLIMVGGGLYLCWGYQMLRGA--LKKEAVSAPAPQVELAKSGRS---FLK 117
Db 61 IETYPALFFTIQLLGAAYLLWLGFAIRSSPPKKEAEIEETQPIQSTKDSKSSIQGFK 120

QY 118 GLLTNLANPKAIYFGVSFLVGDVNGVTTARWGI--FALLIVETLAWFTVVASLFPALPQMR 177
Db 121 GFITNLLNPKALFFLSIFSQFTPQTADWVRWYGLVWVAVGLWFSFLAIFISYKHF 180

QY 178 RGYQRLAKWIDGFAGALFAGFGIHLII 204
Db 181 RPYQTHSYWFDRLFGAALLFAIRIII 207

RESULT 5
B83279
hypothetical protein PA2929 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83279
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID: 20437337
A:Accession: B83279
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004719; GB:AE004091; NID:g9949021; PIDN:AAG06517.1; GSPDB:GN
A:Experimental source: strain PA01
```

```
C:Genetics:
A:Gene: PA2929
C:Superfamily: hypothetical protein b1798

Query Match      26.0%; Score 274.5; DB 2; Length 204;
Best Local Similarity 32.3%; Pred. No. 1.9e-17;
Matches 65; Conservative 42; Mismatches 89; Indels 5; Gaps 2;

QY 2 LMLFLTVAMVHIVALMSPGDPFFVVSOTAVSRKRAMMVGITCGVWVWAGIALLGLH 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 4 LLAFLTTIT--VLAVISPGADFAMVSRNLLYSRRAGLTALGIGAGVTVHVGYSILGV 60
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 62 LIIEKMWLHTLIMVGGGLYLCWMGYQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLIT 121
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 61 VLVRESLALTALKLAGAAYLVFLGLRML--LAREDSVAEEAAGAGVSSWMLRSGLT 118
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 122 NLAMPKALIIYFGVSFLVGDVNGVTARWGIFALIIVETLAWFTVVASLFLPOMRRGYQ 181
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 119 NALNPKTCLFVSLFMQVIDPHTALPAQLGYGAFIALAHVAMFSLVACFLSSPAYRGRLL 178
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 182 BLAKWIDGFAGALFAGFGIHL 202
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 179 RRRRIDQFFGALLVGFVLL 199
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 6
T43921
C:Species: Yersinia enterocolitica
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
C:Accession: T43921
R:Saken, E.M.; Rakin, A.V.; Heesemann, J.
A:Description: Molecular characterization of a novel siderophore-independent iron transport
A:Reference number: Z2724
A:Accession: T43921
A:Status: preliminary; translated from GB/EMBL/DDBY
A:Molecule type: DNA
A:Residues: 1-212 <SAK>
A:Cross-references: EMBL:247200; PIDN:CAB61500.1
A:Experimental source: strain WA-314
C:Genetics:
A:Gene: yfud
C:Superfamily: hypothetical protein b1798
C:Keywords: iron transport

Query Match      25.6%; Score 269.5; DB 2; Length 212;
Best Local Similarity 28.1%; Pred. No. 5.6e-17;
Matches 59; Conservative 52; Mismatches 80; Indels 19; Gaps 4;

QY 3 MLFLTVAMVHIVALMSPGDPFFVVSOTAVSRKRAMMVGITCGVWVWAGIALLGLH 62
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 6 VIILITGL-FVLTFINPGANLLVYVOTSLSSGNAGLTLGLGVAGDAIYSGLGFMAA 64
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 63 IIEKMWLHTLIMVGGGLYLCWMGYQMLR-----GALKKEAVSAPAPQVELAKSGR 113
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 65 LIAEGGALFSAIKIGGLGLVYAYNMYRHRQELHMGMAV--ATSSITPWYF----- 116
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 114 SFLKGLTLNLNPKALIIYFGVSFLVGDVNGVTARWGIFALIIVETLAWFTVVASLFL 173
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 117 -FRRGLLTDLSPQTVLFFTSFVTLPTTPAWAKVMVAMIGIIVASILWRSLLSMAFSR 175
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 174 POMRRGYQRLAKWIDGFAGALFAGFGIHL 203
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 176 PAVRRAYGKVKQHLSSGIGVAVGAFGLRLI 205
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 7
F83444
C:Species: Pseudomonas aeruginosa
C:Superfamily: hypothetical protein PA1620 (imported) - Pseudomonas aeruginosa (strain PA01)
```

```
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83444
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.D.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: F83444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-213 <STO>
A:Cross-references: GB:AE004589; GB:AE004754; PIDN:AAG05009.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1620
C:Superfamily: hypothetical protein b1798

Query Match      25.0%; Score 263.5; DB 2; Length 213;
Best Local Similarity 27.3%; Pred. No. 1.9e-16;
Matches 59; Conservative 46; Mismatches 92; Indels 19; Gaps 3;

QY 2 LMLFLTVAMVHIVALMSPGDPFFVVSOTAVSRKRAMMVGITCGVWVWAGIALLGLH 61
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 4 LSVLLTAAVFAVALVSPGPDVALVVRTSLHGRAGLASALGLACGILLHTLTLVTGVS 63
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 62 LIIEKMWLHTLIMVGGGLYLCWMGYQMLRGALKK-----EAVSAPAPQVELAKSGR 113
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 64 LLLSRTPVFLFATLQALGALYLAWLVGALRAWLRGDSQPGRLDGLALPPSP-----G 116
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 114 SFLKGLTLNLNPKALIIYFGVSFLVGDVNGVTARWGIFALIIVETLAWFTVVASLFL 173
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 117 PWRGVTATNLPKALVPFFIALLGSLIPQAQMSLGGKLAIVAAALLFGMACWFLSLTLTR 176
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 174 POMRRGYQRLAKWIDGFAGALF----AGFGIHLIIS 205
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 177 PALQARLLRAVPWLDACGVVFLVAAAIIHLVRS 212
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 8
H64759
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: H64759
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H64759
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-223 <BLAT>
A:Cross-references: GB:AE000140; GB:U00096; PIDN:g1786520; PIDN:AAC73431.1; PID:g17865
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: yahN
C:Superfamily: hypothetical protein b1798
C:Keywords: transmembrane protein
F:26-42/Domain: transmembrane #status predicted <TM1>
F:85-101/Domain: transmembrane #status predicted <TM2>
F:163-179/Domain: transmembrane #status predicted <TM3>
F:200-216/Domain: transmembrane #status predicted <TM4>

Query Match      24.8%; Score 261.5; DB 2; Length 223;
Best Local Similarity 28.2%; Pred. No. 3e-16;
Matches 58; Conservative 50; Mismatches 83; Indels 15; Gaps 3;

QY 4 LFLTVAMVHIVALMSPGDPFFVVSOTAVSRKRAMMVGITCGVWVWAGIALLGLH 63
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

```
Db 20 VYLTVGL-FVITTFNPGANLFVVVQTSLSAGRRAGVLTGLGVALGDADFVSGIGLGLATL 78
Qy 64 IEKMWLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPOVELAKSGRS-----FLK 117
Db 79 ITOCEEISLIRIVGAILWFAFNCMR-----RQSTPQMSITLOQPIAPWVVFPR 130
Qy 118 GLITNLNPKKAIYFGSVFSLFVGDNDVGTARWGIIFALIIVETLAWFTVVASLFAIPQMR 177
Db 131 GLITDLSNQTVLFFISIFSVTLNATPTWRLMAMAGIVLASIIMRVFSLQASLPAVR 190
Qy 178 RGFORLAKWIDGFAGALFAGFGIHLI 203
Db 191 RAYGRMORVASRVIGAIIGVFAIRLI 216

RESULT 9
F83306
hypothetical protein PA2710 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83306
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: A82950; MUID:20437337
A:Accession: F83306
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004699; GB:AE004091; NID:99948782; PIDN:AAG06098.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
C:Superfamily: hypothetical protein bl798

Query Match 23.4%; Score 246.5; DB 2; Length 204;
Best Local Similarity 23.2%; Pred. No. 6e-15;
Matches 62; Conservative 35; Mismatches 100; Indels 15; Gaps 2;

Qy 1 MLMFLTAMVHIVAMSPGDPFFVFSQTAVSRSKRAMMVGILTCGVMYWGATLGL 60
Db 1 MLFTSLFVATLTCMLSPGDPDFLLIRNAARYORSAMMTSLGVLGVATHMAYCVAGL 60
Qy 61 HLIIERKMWLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPOVELAKSG-----R 113
Db 61 AVLIITTPWLFNALXYTGAVLIWIGIQALR-----SRGGTLDLAVGVGVGVGHS 112
Qy 114 SFLLKGLITNLNPKKAIYFGSVFSLFVGDNDVGTARWGIIFALIIVETLAWFTVVASLFA 173
Db 113 AFLQGLICLLNLPKATLFLAVFTQVILSDLOSSFAELKTAGIIVGLAVLWMLPVLVLIQS 172
Qy 174 PQMRGQRORLAKWIDGFAGALFAGFGIHLIS 205
Db 173 AVVRSRLARAGGVVDKLLGGLLIALGVKVALS 204

RESULT 10
G83082
hypothetical protein PA4507 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83082
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: A82950; MUID:20437337
A:Accession: G83082
A:Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AE004864; GB:AE004091; NID:99950740; PIDN:ARG07895.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
C:Superfamily: hypothetical protein bl798

Query Match 23.2%; Score 245; DB 2; Length 210;
Best Local Similarity 33.3%; Pred. No. 8.4e-15;
Matches 75; Conservative 34; Mismatches 58; Indels 48; Gaps 12;

Qy 6 LTVAMVHIVAMSPGDPFFVFSQTAVSRSKRAMMVGILTCGVMYWGATLGL 61
Db 8 TALAVYLV-----PGDMLLLFOTGARQGRRAALVTALGALARACHVL-----NAATGLA 59
Qy 62 LIIEKMWLHTLIMVGGGLYLCWMGYOMLRG---AL-KKEAVSAP-APQVELAKSGRSFL 116
Db 60 LLFERTAPWTFDLVRLIGAVYLAWLQWLGRGGGLALFTSDAGSAPVPHAD-----RRALL 115
Qy 117 KGLITNLNPKKAIYFGSVFSLFVGDNDVGTARWGIIFALIIVETLAWFTVVASL----- 170
Db 116 RGLITNLNPKKALFCSVLLPQFVSPKAGSLA-----VQFAALGTVLVLVCLAFDC 166
Qy 171 -PALPQMRG-----YORLAKWIDGFAGALFAGFGIHLIS 206
Db 167 AYALAGRLGRWLASRPRAQLQW--GF-GGLLIGFVRLALLR 208

RESULT 11
C82979
hypothetical protein PA5341 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C82979
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: C82979
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AE004946; GB:AE004091; NID:99951650; PIDN:ARG08726.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
C:Superfamily: hypothetical protein bl798

Query Match 19.8%; Score 208.5; DB 2; Length 206;
Best Local Similarity 28.8%; Pred. No. 1.5e-11;
Matches 65; Conservative 32; Mismatches 86; Indels 43; Gaps 9;

Qy 1 MLMFLTAMVHIVAMSPGDPFFVFSQTAVSRSKRAMMVGILTCGVMYWGATLGL 60
Db 1 MEAFAYLVASTHFAALLSPGDPDFLLVRAALLRRRQAD---GVAAGI---ALANLLSM 53
Qy 61 HLIIERKMA-----WLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPOVELA 109
Db 54 LVLVGLLASVPDSAHSAURLQAL---CGLYFLWLAGOAL---LAQRLEMPAQRDVPS 106
Qy 110 KSG--RSFLKGLITNLNPKKAIYFGSVFSLFVGDNDVGTAR-----WGI---FALIVE 159
Db 107 RGVLRGLRDGLASSLNPKLPIFYAGLF-----GVLARFSLPGNALCLAWNSLA 158
Qy 160 TLAWFTVVASLFAIPQMRGQRORLAKWIDGFAGALFAGFGIHLIS 205
Db 159 VLCWDNALIVRLDRPRWGVQRVVGALDRLCGVLLALLGGWLVLA 204
```

```
RESULT 12
P83051
conserved hypothetical protein PA4757 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83051
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: F83051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <STO>
A:Cross-references: GB:AE004889; GB:AE004091; NID:g9951014; PIDN:AAG08143.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4757
C:Superfamily: hypothetical protein bl798

Query Match 19.6%; Score 207; DB 2; Length 216;
Best Local Similarity 29.4%; Pred. No. 2.1e-11;
Matches 64; Conservative 30; Mismatches 86; Indels 36; Gaps 4;

QY 5 FLTVAMVHVIALMSGPGDPFFVTSQTAVSRSRKEAMMGVIGITCGVWVWAGIALGLHLII 64
DB 9 FWTYVLGVVVFVILLPGPNSLFLVATSAQRGVATGYRAACGVFLGDAVLMLLSALGVASLL 68

QY 65 EKMAWLHTLMVGGGLYLCWGYOMLRGALKK----EAVSAPAPQVELAKSGRSFLKGL 120
DB 69 KAEPMLFGLKYLGAAYLFYLGVMGLRGAWKRLNPEATAAQAQVDV---HQPFRKALL 125

QY 121 TNLANKPAIIYGSVFSFVSGDNGVTTARMGIFALIIVETL-----AWF 164
DB 126 LSLSNPKALFFISFFIQFVDPGYAYPGLSFLVLAIVLELSALYLSFLIFTGVRLAAMF 185

QY 165 TVVASLFPALPQMRGQYRLAKWIDGAFAGLFGFIHL 202
DB 186 -----RRQRLLAGATSGVGLFVGFKL 210

RESULT 13
E83703
hypothetical protein BH0429 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83703
R:Takami, R.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: E83703
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BA004148.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0429

Query Match 19.1%; Score 201; DB 2; Length 207;
Best Local Similarity 28.3%; Pred. No. 7e-11;
Matches 63; Conservative 39; Mismatches 81; Indels 40; Gaps 9;

QY 2 LMLPLTVAMVHVIALMSGPGDPFFVTSQTAVSRSRKEAMMGVIGITCGVWVWAGIALGLH 61
DB 6 LLSFLGVA---VLLTLMGPDILFLVAQSMNQRAQIVTALGLCTGLLVHITAAVVGIS 62
```

```
QY 62 LIIEKMAWLHTLMVGGG---LYLCWGYOMLRGALKKEAVSAPAPQVELAKSGRSFLKGL 118
DB 63 ALIYOSALAFVTVKAYAGAVLLYLAWKAPQE-----KGEGLSID-KOTTIA-YGALYKKG 115

QY 119 LUTNLANKPAIIYGSVFSFVSGDNGVTTARMG-----IFALIIVETLAWFT---- 165
DB 116 IIMNVLNPKVSLFFLALLPQFVNSGAG-SAPQWMLLLGCVFVLQAFIIFSLVSWFAEKVG 174

QY 166 --VVASLFPALPQMRGQYRLAKWIDGAFAGLFGFIHLISR 206
DB 175 QLLMRSSPIKQMR-----IRGGLLALIGLQVAFSK 206

RESULT 14
C55580
hypothetical protein (carA 3' region) - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 29-Sep-1999
C:Accession: C55580
R:Kwon, D.H.; Lu, C.D.; Walthall, D.A.; Brown, T.M.; Houghton, J.E.; Abdelal, A.T.
J. Bacteriol. 176, 2532-2542, 1994
A:Title: Structure and regulation of the carAB operon in Pseudomonas aeruginosa and P
A:Reference number: A55580; MUID:94222830
A:Accession: C55580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <KNO>
A:Cross-references: GB:004992; NID:g451649; PIDN:AAA19047.1; PID:g451652
C:Superfamily: hypothetical protein bl798

Query Match 18.9%; Score 199; DB 2; Length 216;
Best Local Similarity 28.4%; Pred. No. 1.1e-10;
Matches 62; Conservative 31; Mismatches 89; Indels 36; Gaps 4;

QY 5 FLTVAMVHVIALMSGPGDPFFVTSQTAVSRSRKEAMMGVIGITCGVWVWAGIALGLHLII 64
DB 9 FWTYVLGVVVFVILLPGPNSLFLVATSAQRGVATGYRAACGVFLGDAVLMLLSALGVASLL 68

QY 65 EKMAWLHTLMVGGGLYLCWGYOMLRGALKK----EAVSAPAPQVELAKSGRSFLKGL 120
DB 69 KAEPMLFGLKYLGAAYLFYLGVMGLRGAWKRLNPEATAAQAQVDV---HQPFRQALL 125

QY 121 TNLANKPAIIYGSVFSFVSGDNGVTTARMGIFALIIVETL-----AWF 164
DB 126 LSLSNPKALFFISFFIQFVDPGYAYPGLSFLVLAIVLELSALYLSFLIFTGVRLAAMF 185

QY 165 TVVASLFPALPQMRGQYRLAKWIDGAFAGLFGFIHL 202
DB 186 -----RRQRLLAGATSGVGLFVGFKL 210

RESULT 15
B82410
conserved hypothetical protein VCA0846 [imported] - Vibrio cholerae (group O1 strain
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: B82410
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82410
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <HEI>
A:Cross-references: GB:AE004412; GB:AE003853; NID:g9658269; PIDN:AAF96744.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16361; biotype El Tor
C:Genetics:
A:Gene: VCA0846
A:Map position: 2
```

Query Match 18.7%; Score 197.5; DB 2; Length 204;
Best Local Similarity 25.2%; Pred. No. 1.4e-10;
Matches 51; Conservative 46; Mismatches 102; Indels 3; Gaps 1;

QY 2 LMLRLTVMVHIYALKSPGDFEFYSOTAVSRKRKEMMGVLGITCGVMVWAGIALLLGLH 61
DB 3 LTVWLSLFTICILGAMSPGSLANWAKHSLAGGRKNGFAAAWAHAFGIGVYAFITLIGLA 62

QY 62 LIIEKMAWLHTLIMVGGGLYLCWMGYQMLRGALKKEAVSAPQVELAKSGRSLKGLLT 121
DB 63 VVLHQSPVLFRTISYAGAAYLAYLGNALR---SKGGVAAKLESGESVSNWQSAREGLLI 119

QY 122 NLAMPKAIYFGSVFSLFVGDNVGTARWGIPALIIVETLAWFTTWVASLFPALPQMRGQY 181
DB 120 SLLSPKIALFFIALFSQYVANGSDLTSKAAIVITPLVVDGLWYSFITILILSSPRLLOKLR 179

QY 182 RLAKWIDGFAGALFAGFGIHLI 203
DB 180 ARAVLIDRLSLGLVLAIRVL 201

Search completed: May 6, 2001, 14:40:02
Job time: 343 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:41:20 ; Search time 16.79 Seconds
(without alignments)
420.288 Million cell updates/sec

Title: US-09-466-935-4

Perfect score: 1054

Sequence: 1 MLMLFTVAWVHVALMSPG.....IDGFAGALFAGFGIHLIISR 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1054	100.0	206	1 RHTC_ECOLI	P27846 escherichia
2	368.5	35.0	210	1 YD07_HAEBIN	P57320 haemophilus
3	261.5	24.8	223	1 YAHN_ECOLI	P75693 escherichia
4	207	19.6	216	1 YBF7_PSEAE	P38102 pseudomonas
5	190.5	18.0	212	1 YEAS_ECOLI	P76219 escherichia
6	180.5	17.1	210	1 YH2P_BACSU	O05406 bacillus su
7	159	15.1	203	1 CHPE_PSEAE	O87005 pseudomonas
8	144.5	15.1	206	1 YG27_STNY3	P74343 synecocyst
9	144.5	13.7	206	1 RHTB_ECOLI	P27847 escherichia
10	111.5	10.6	206	1 YGGA_AERSA	P70775 aeromonas s
11	105	10.0	236	1 LYSE_CORGL	P94633 corynebacte
12	104.5	9.9	225	1 YGGA_AERHY	P52047 aeromonas h
13	103	9.8	211	1 YGGA_ECOLI	P1667 escherichia
14	101	9.6	195	1 YF1K_ECOLI	P38101 escherichia
15	96.5	9.2	201	1 Y488_MYCTU	Q11154 mycobacteri
16	92	8.7	405	1 WCAD_ECOLI	P71238 escherichia
17	89	8.4	199	1 YJ86_MYCTU	Q10871 mycobacteri
18	88	8.3	541	1 COX1_BRAJA	P31833 bradyrhizob
19	87.5	8.3	722	1 VATI_HALN1	Q9HND8 halobacteri
20	85.5	8.1	695	1 DIP2_MYCTU	Q10801 mycobacteri
21	84.5	8.0	459	1 NU4M_POLOR	Q95917 polypterus
22	84.5	8.0	663	1 CYOB_ECOLI	P18401 escherichia
23	83.5	7.9	193	1 LSP2_VITS1	O9K450 vitreoscilli
24	83	7.9	464	1 GNP1_ECOLI	P37021 escherichia
25	82.5	7.8	653	1 YTH3_CABEL	P34196 crossstoma
26	82	7.8	216	1 RIB7_MEITH	O26337 methanobact
27	81	7.7	173	1 NU6M_CHOLA	Q49131 methylobact
28	80.5	7.6	287	1 MAGN_METEX	P31125 escherichia
29	80.5	7.6	299	1 EAMA_ECOLI	P31125 escherichia
30	80.5	7.6	299	1 YDED_ECOLI	P31125 escherichia
31	80.5	7.6	1080	1 CYAT_HUMAN	P51828 homo sapien
32	80.5	7.6	607	1 NUSM_CARAU	O78688 carassius a
33	80				

34	79.5	7.5	379	1	CYB_AKOB0	P21714 akodon boli
35	79.5	7.5	379	1	CYB_DASNO	O21337 dasypus nov
36	79.5	7.5	459	1	NU4M_HALGR	P38601 halichoerus
37	79	7.5	172	1	NU6M_CYPCA	P24982 cyprinus ca
38	79	7.5	452	1	NU4M_BRAFL	O47423 branchiost
39	79	7.5	452	1	NU4M_BRAFL	O79421 branchiost
40	78.5	7.4	133	1	PTPW_ECOLI	P42905 escherichia
41	78.5	7.4	173	1	NU6M_CARAU	O78689 carassius a
42	78.5	7.4	662	1	CYOB_BOCAL	P57543 bucinera ap
43	78	7.4	493	1	SECY_ARCFU	O28377 archaeglob
44	78	7.4	534	1	COX1_KUGLA	P20386 kluyveromyc
45	77.5	7.4	262	1	COX3_METSE	O47491 metridium s

ALIGNMENTS

RESULT 1
RHTC_ECOLI
ID RHTC_ECOLI STANDARD; PRT; 206 AA.
AC P27846;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE THREONINE EFFLUX PROTEIN.
GN RHTC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
RN [2]
RP REVISIONS.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE OF 1-107 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=87115164; PubMed=3027506;
RA Irino N., Nakayama K., Nakayama H.;
RT "The recQ gene of Escherichia coli K12: primary structure and
evidence for SOS regulation."
RL Mol. Gen. Genet. 205:298-304(1986).
RN [4]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RT Unpublished observations (DEC-1997).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=99313167; PubMed=10386596;
RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
Livshits V.A.;
RT "The novel transmembrane Escherichia coli proteins involved in the
amino acid efflux."
RL FEBS Lett. 452:228-232(1999).
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
INTRODUCED IN POSITION 80 TO PRODUCE THIS ORF.

```

CC CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC CC IN POSITIONS 47 AND 73.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M87049; AAA67619.1; ALT.FRAME.
DR EMBL; AE000458; AAC76826.1; ALT.FRAME.
DR EMBL; M30198; -. NOT_ANNOTATED_CDS.
DR PIR; S30713; S30713.
DR EcoGene; EG11468; rhtC.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; LyseB; 1.
KW Transport; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 67 87 POTENTIAL.
FT TRANSMEM 150 173 POTENTIAL.
SQ SEQUENCE 206 AA; 22474 MW; F64017878CC6D50D CRC64;

Query Match 100.0%; Score 1054; DB 1; Length 206;
Best Local Similarity 100.0%; Pred. No. 2.9e-77;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLFLTVAMVHIVAMSPGDPFFVSVTSRKRKAMGMVLGTCGVVWVWAGIALGL 60
DB 1 MMLFLTVAMVHIVAMSPGDPFFVSVTSRKRKAMGMVLGTCGVVWVWAGIALGL 60
QY 61 HLIEKMAWHLTLWGGGLYLCWGYOMLKGALKKEAVSAPAPQVELAKSRFLKGL 120
DB 61 HLIEKMAWHLTLWGGGLYLCWGYOMLKGALKKEAVSAPAPQVELAKSRFLKGL 120
QY 121 TNLANPKAIYFGSVFSLFVGDNVGTARWGIPALIIIVETLAWFTVVASLPALQMRRGY 180
DB 121 TNLANPKAIYFGSVFSLFVGDNVGTARWGIPALIIIVETLAWFTVVASLPALQMRRGY 180
QY 181 ORLAKWIDGFAAGLAFAGFGIHLIISR 206
DB 181 ORLAKWIDGFAAGLAFAGFGIHLIISR 206

RESULT 3
YD07_HAETN
ID YD07_HAETN STANDARD; PRT; 210 AA.
AC Q57320; Q05057;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL PROTEIN H11307.
GN H11307.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;

```

```

RT *whole-genome random sequencing and assembly of Haemophilus
RT Influenzae Rd.;
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32810; AAC22954.1; -.
DR TIGR; H11307; -.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; LyseB; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 42 62 POTENTIAL.
FT TRANSMEM 86 86 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
SQ SEQUENCE 210 AA; 23636 MW; 254D159014845473 CRC64;

Query Match 35.0%; Score 368.5; DB 1; Length 210;
Best Local Similarity 38.8%; Pred. No. 8.8e-23;
Matches 80; Conservative 42; Mismatches 73; Indels 11; Gaps 4;

QY 4 LEFTVAMVHIVAMSPGDPFFVSVTSRKRKAMGMVLGTCGVVWVWAGIALGLHLI 63
DB 1 MMLFLTVAMVHIVAMSPGDPFFVSVTSRKRKAMGMVLGTCGVVWVWAGIALGL 60
QY 64 IEKMAWHLTLWGGGLYLCWGYOMLKGALKKEAVSAPAPQVELAKS---GRSLKGL 120
DB 61 FVTIPALHGVYMLGGSLAYLGLFLMARS--KYAKFESHSDTEFNQQTIKKELKGL 118
QY 121 TNLANPKAIYFGSVFSLFVGDNVGTARWGIPALIIIVETLAWFTVVASLPALQMRRGY 177
DB 119 VNTSNKRVVYVYSSVMSLVL---VNITEMMOILAFAVIVVETFCYFYVILFISRTAK 175
QY 178 RGYORLAKWIDGFAAGLAFAGFGIHLI 203
DB 176 RLYSQVSRVIDNMAGIVFLFGCVLV 201

RESULT 3
YAHN_ECOLI
ID YAHN_ECOLI STANDARD; PRT; 223 AA.
AC P75693; P71307;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 24.8 KDA PROTEIN IN BETT-PRPR INTERGENIC REGION.
GN YAHN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MC1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.

```

```

RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RL -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000140; AAC73431.1; -
DR EMBL: U73857; AAB18053.1; -
DR ECGene: EG13598; yahn.
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 213 223 LRLYEGVTOR -> YA (IN REF. 2).
SQ SEQUENCE 223 AA; 24811 MW; E17F5ABC31EE3F26 CRC64;

Query Match 24.8%; Score 261.5; DB 1; Length 223;
Best Local Similarity 28.2%; Pred. No. 3e-14;
Matches 58; Conservative 50; Mismatches 83; Indels 15; Gaps 3;

QY 4 LFLVAMVHVALMSPGDPFFVSVTSVSRKEMMGVLGTCGVMMWAGIALGLHLI 63
:||||: ::|||: |||||: ::|||: |||||: |||||: |||||: |||||:
DB 20 VLTAVGL-FVITFENPGANLFFVVTQTSLSGRAGVLTGLGVALGDAFYSGIGLGLATL 78
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 64 LEKMAWLTLMVGGGLYLCWGYQMLRGALKKAVSAPQVELAKGSR-----FLK 117
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 79 ITQCEIFSLIRIVGGAYLLNFAWCSMR-----RQSTPQMSTLQQPISAPWVYFFRR 130
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 118 GLLTNLNPKAIYFGSVFLVGVNDVGTARWGIFALIVETLAWFTVVASLFAIPQMR 177
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 131 GLITDLSNPQTVLFFISFVTLNAETPTWRLMAWAGIVLAIIRWFVLSQATSLPAVR 190
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 178 RYORLAKWIDGAGALPAGFGIHLI 203
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 191 RAYGMORVASRVIGALIGVFARLI 216
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

RESULT 4
YBF7_PSEAE STANDARD; PRT; 215 AA.
AC P38102; Q9HV45;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4757.
GN PA4757.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RC MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U04992; AAA19047.1; -
DR EMBL: U81259; AAB39251.1; -
DR EMBL: AE000489; AAG08143.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT CONFLICT 109 109 A -> G (IN REF. 1).
FT CONFLICT 119 119 P -> R (IN REF. 1).
FT CONFLICT 122 122 K -> Q (IN REF. 1).
SQ SEQUENCE 216 AA; 23248 MW; A30A0BE714591B8D CRC64;

Query Match 19.6%; Score 207; DB 1; Length 216;
Best Local Similarity 29.4%; Pred. No. 6.2e-10;
Matches 64; Conservative 30; Mismatches 88; Indels 36; Gaps 4;

QY 5 ELTVAMVHVALMSPGDPFFVSVTSVSRKEMMGVLGTCGVMMWAGIALGLHLI 64
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 9 FWTYVLGVVFIILPGPNSLFLVLTSAQRGATGYRACGVFLGDAVLMLLSALGVASIL 68
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 65 EKMAWLTLMVGGGLYLCWGYQMLRGALKK-----EAVSAPAPQVELAKSGRSLKGLL 120
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 69 KAEPMLFGLKALYALFLYLVGMGRGAWKRLNPNPEATAQAQEDV---HQPFRKALL 125
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 121 TNLANKAIYFGSVFLVGVNDVGTARWGIFALIVETL-----AMF 164
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 126 LLSNPKAILFFISFFIQFVDPGYAYPGLSFLVLAIVLELVSALYSFLITGVRLAAWF 185
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
QY 165 TVVASLFAIPQMRGYQRLAKWIDGAGALPAGFGIHL 202
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
DB 186 -----RRQRLAAGATSGVGLFVFGVKL 210
:||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||:

RESULT 5
YEAS_ECOLI STANDARD; PRT; 212 AA.
ID YEAS_ECOLI
AC P76249; O07971; O07969;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 23.2 KDA PROTEIN IN GAPA-RND INTERGENIC REGION.
GN YEAS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```
OC Escherichia.
OX NCBI_TaxID=562;
RN
RP
RX MEDLINE=97426617; MG1655;
RC STRAIN=K12;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Motomura K., Nakade S., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996)
RC
RL
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000274; AAC74868.1; -
CC EMBL; D90823; BAA15593.1; -
CC EMBL; D90824; BAA15602.1; -
CC EcoGene; EG13505; Yeas.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; Lyse; 1.
CC KW Hypothetical protein; Transmembrane.
CC TRANSMEM 12 32 POTENTIAL.
CC TRANSMEM 49 69 POTENTIAL.
CC TRANSMEM 71 91 POTENTIAL.
CC TRANSMEM 122 142 POTENTIAL.
CC TRANSMEM 153 173 POTENTIAL.
CC TRANSMEM 188 208 POTENTIAL.
CC SEQUENCE 212 AA; 23200 MW; 342E0DF348C9AD9A CRC64;
SQ
Query Match 18.0%; Score 190; DB 1; Length 212;
Best Local Similarity 28.2%; Pred. No. 1,4e-08;
Matches 57; Conservative 36; Mismatches 85; Indels 24; Gaps 6;
QY 13 IVALMSPGDFFVVSQTAVERSRKEAMWGVLITCG-----VWVAGIALGLHLIEKM 67
DB 18 IFIVLPGPNTLVKNSVSGMKGGYLACGVFIGDAVLMFLAWAGVATL-----IKTT 72
QY 68 AMLHTLIMVGGGLYLCWGYOMLRGALKKEAVSAPAPQVEVLAKEGRSFLKGLTJNLNPK 127
DB 73 PLFLNIVRYLGAFLYLLGSKILYATLKGNSEAKSDPEQV---GAIFKRALILSLTNPK 129
QY 128 AIIYEGSVFSFLVGDVNGVTTARWGIFALIVETLAW-----FTVVASLFLPQMRRGY 180
DB 130 ALIFKVSFFVQFIDVNAFHT---GISFILAATLELVSPCVLSFLIISGAF-VTQYIRTK 185
QY 181 ORLAKWIDGCFAGALFAGFGHIL 202
DB 186 KKLAKVGNLSLGLMFGVGAARL 207
RESULT 6
YRHP-BACSV STANDARD; PRT; 210 AA.
ID
AC 095406; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE HYDROTHERICAL 23.4 KDA PROTEIN IN AAPA-SIGV INTERGENIC REGION.
GN YRHP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN
RP
RX MEDLINE=97453479; PubMed=9308178;
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RA Duesterhoef A., Ehrlich S.D.;
RT "Sequence of the Bacillus subtilis genome region in the vicinity of
RT the lev operon reveals two new extracytoplasmic function RNA
RT polymerase sigma factors Sigv and Sigz.";
RL Microbiology 143:2939-2943(1997)
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z99117; CAB14652.1; -
CC EMBL; U93874; AAB80873.1; -
CC Subtilist; BG12304; Yrhp.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; Lyse; 1.
CC KW Hypothetical protein; Transmembrane.
CC TRANSMEM 5 25 POTENTIAL.
CC TRANSMEM 50 70 POTENTIAL.
CC TRANSMEM 75 95 POTENTIAL.
CC TRANSMEM 155 175 POTENTIAL.
CC SEQUENCE 210 AA; 23389 MW; 2A37D9419FDB0A58 CRC64;
SQ
Query Match 17.1%; Score 180.5; DB 1; Length 210;
Best Local Similarity 24.5%; Pred. No. 7,7e-08;
Matches 51; Conservative 41; Mismatches 101; Indels 15; Gaps 5;
QY 2 LMLFLTVAMVHIVALMSPGDDFFVVSQTAVERSRKEAMWGVLITCGVWVWAGIALGLH 61
DB 4 LLAIPIDAMWVIL---PGADTLMVKNLTVRGPRAGRNILGLATGLSPWTVAILGLS 60
QY 62 LIERKAMVHLIMVGGGLYLCWGMGTOMLRGALKKEAVSAPAPQVE---LAKSGR----- 113
DB 61 VVIKSVILFTTKIKYLGAAALYILG---VKSFSAKSMFLSDDDMSQAKNWSKPKRYKT 117
QY 114 SFLKGLTLNLANPKAIIYFGSVSLFVGDVNGVTTARWGIFALII-VETLAWFTVVASLFA 172
DB 118 SPWQSLSNILNPKTVLVVYVYTIMPQFINLNGINQQLILASILTLLAVLWFLFLYIID 177
QY 173 LPQMRRGYORLAKWIDGCFAGALFAGFGI 200
DB 178 YAKKWKNSKFKQVFKITGITLILVGFGI 205
```

```
RESULT 7
CHPE_PSEAE
ID CHPE_PSEAE STANDARD; PRT; 203 AA.
AC O87005;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHEMOTACTIC TRANSDUCTION PROTEIN CHPE.
GN CHPE OR PA0417.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
ON NCBI_TaxID=287;
RX [1]
SEQUENCE FROM N.A.
RC STRAIN=PA01;
RA "Pseudomonas aeruginosa chemotactic transduction genes pill, chpA
RT chpB and downstream genes chpC, chpD and chpE.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gaber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964 (2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U79580; AAC23935.1; -
CC EMBL; AE004479; AAG03806.1; -
CC InterPro: IPR001123; -
CC Pfam: PF01810; LYSE; 1.
CC Transmembrane.
CC TRANSMEM 3 23 POTENTIAL.
CC TRANSMEM 45 66 POTENTIAL.
CC TRANSMEM 59 89 POTENTIAL.
CC TRANSMEM 123 143 POTENTIAL.
CC TRANSMEM 149 169 POTENTIAL.
CC SEQUENCE 203 AA; 21290 MW; 195553C048AAD099 CRC64;
Query Match 15.1%; Score 159; DB 1; Length 203;
Best Local Similarity 27.7%; Pred. No. 3.8e-06;
Matches 59; Conservative 38; Mismatches 86; Indels 30; Gaps 8;
QY 1 MMLFLTVAMVHIVALMSFGDPFFVSQTAVSRKKEAMMGVLGITCGYVWVWAGIALGL 60
DB 1 MLIFFLAALLFGFAFNVSPGAVFSETLRGTGGFRPALVQLGSLIGDANVWLLGLTGL 60
QY 61 HLII--EKMAWLHTLMVGGGLYLCWQYQMLRGALKKEAVSAPAPQVELAKSGR-SFLK 117
DB 61 ALLLGEQVRIPUTLACAA--YLAWLGVQGLR-----DAWSPPLAEDAGEGQGNAPGA 112
QY 116 GLLTNLANPKAIIFYGVSFLVGNVGT---TARWGIFALIIVETLAWFTVVASLFPALP 174
DB 113 GAATSLSNPNVYWGALGSLAGALGIVDCTPNQAQSLVFFAGFMLSLWCFCCRAL--VD 170
```

```
QY 175 QMRG-----YQPLAKWIDGPAALFAGFGIHLI 203
DB 171 WLARNTSLFWHRVS-----YAGCGVLLL 193
RESULT 8
YD27_SYNY3
ID YD27_SYNY3 STANDARD; PRT; 206 AA.
AC P74343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 22.0 KDA PROTEIN SLR1627.
GN SLR1627.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
ON NCBI_TaxID=1148;
RX [1]
SEQUENCE FROM N.A.
MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Rosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Tamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D90914; BAA18437.1; -
CC InterPro: IPR001123; -
CC Pfam: PF01810; LYSE; 1.
CC Hypothetical protein; Transmembrane.
CC TRANSMEM 9 29 POTENTIAL.
CC TRANSMEM 47 67 POTENTIAL.
CC TRANSMEM 74 94 POTENTIAL.
CC TRANSMEM 150 170 POTENTIAL.
CC SEQUENCE 206 AA; 22041 MW; EF493754B8F264AF CRC64;
Query Match 15.1%; Score 159; DB 1; Length 206;
Best Local Similarity 26.8%; Pred. No. 3.9e-06;
Matches 56; Conservative 39; Mismatches 96; Indels 18; Gaps 6;
QY 3 MFLFTVMVHIVALMSFGDPFFVSQTAVSRKKEAMMGVLGITCGYVWVWAGIALGLHL 62
DB 9 ILSLFAMLIILAAAL--PSSLVLTVSSKSASGGFIHGLFAALGVVLGDIIFILIALWGLAF 66
QY 63 IEKMAWLHTLMVGGGLYLCWQYQMLRGALKKEAVSAPAPQVELAKSGRSLKGLTN 122
DB 67 LEGAMGDFEVILKVTISGYLSWLGIINTIRAKVNNQSLA----KVDVKSLSFSFAGLIT 122
QY 123 LANPKAIIFYGVSFLVGNVGTARWGIFALIIVETLAWFTVVASLFPALPQMRG-- 180
DB 123 LADQKAVFLYGLFPTFV--DVNNIAVLDI-AVILITAILTVGGVKIFYAFLAHRSGLLI 179
QY 181 ----ORLAKWIDGPAALFAGFGIHLIIS 205
DB 180 SRQNKRIINYL---AGALMISVGVLLIS 205
RESULT 9
```

```
RHTB_ECOLI
ID RHTB_ECOLI STANDARD; PRT; 206 AA.
AC P27847;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMOSERINE/HOMOSERINE LACTONE EFFLUX PROTEIN.
GN RHTB
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RL from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JUN-1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=99313167; PubMed=10386596;
RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
RA Livshits V.A.;
RT "The novel transmembrane Escherichia coli proteins involved in the
RT amino acid efflux";
RL FEBS Lett 452:228-232(1999).
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF HOMOSERINE AND HOMOSERINE
CC LACTONE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
CC BE INTRODUCED IN POSITIONS 60 TO EXTEND THIS ORF.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M87049; AAA67620.1; ALT_FRAME.
CC PIR; S30714; S30714.
CC EcoGene; EGI1469; rhtB.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; Lyse; 1.
CC Transport; Transmembrane.
CC FT TRANSMEM 5 25 POTENTIAL.
CC FT TRANSMEM 45 65 POTENTIAL.
CC FT TRANSMEM 68 88 POTENTIAL.
CC FT TRANSMEM 117 137 POTENTIAL.
CC FT TRANSMEM 148 168 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC SEQUENCE 206 AA; 22427 MW; 1164F17738509C8C CRC64;

Query Match 13.78; Score 144.5; DB 1; Length 206;
Best Local Similarity 20.78; Pred. NO. 5.5e-05;
Matches 43; Conservative 44; Mismatches 96; Indels 25; Gaps 4;

QY 10 MVHIVALSPGDPFFVSQTAVSRKXKAMGVLTGTCVMWAGIALGLHLIERKAW 69
D 1 L7SITLSLSPGSGAINTWTTSLNHYRGAVASTAGTQTLAIHIVGVGLGTLFSRSVI 70
QY 70 LHTLIMVGGGLYLCWMGYOMLR--GALKKEAVSAPQVELAKSGRSFLKGLLTNLANPK 127
```

```
Db 71 AFEVLKWAAGAAVLIWLGTOQWRAAGAIKLSLASTQSRRL-----FQRAVFNLTNPK 124
QY 128 AIIYFGVSFLFVGDVNGTGTARWGIFALIIVETLAWFTVVASLFAIPOMRRGYORLAKWI 187
D 125 SIVFLAALFPQFIPQOPOLMOYIVLGVT-----TIVDIIWIMGYATLAQRIALWI 176
QY 188 DG-----PAGALFAGFGIHLIISR 206
D 177 KGPQKMKALNKIFGSLFMLVGAALLASAR 204

RESULT 10
YGGA_AERSA STANDARD; PRT; 206 AA.
AC P70775;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 21.5 KDA PROTEIN IN ASAR-CDPD INTERGENIC REGION.
GN YGGA.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIMB 1102;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RA Williams P., MacIntyre S., Stewart G.S.A.B.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U65741; AA070019.1; ALT_INIT.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; Lyse; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 37 57 POTENTIAL.
CC FT TRANSMEM 65 85 POTENTIAL.
CC FT TRANSMEM 116 136 POTENTIAL.
CC FT TRANSMEM 148 168 POTENTIAL.
CC FT TRANSMEM 185 205 POTENTIAL.
CC SEQUENCE 206 AA; 21505 MW; D1C2C492CDA0179A CRC64;

Query Match 10.68; Score 111.5; DB 1; Length 206;
Best Local Similarity 26.08; Pred. NO. 0.023;
Matches 52; Conservative 32; Mismatches 91; Indels 25; Gaps 11;

QY 14 VALMSP-GDPFFVSQTAVSRKXKAMGVLTGTC-----GYVWAGIALGLHLI-IEK 66
D 13 LAMITPTGAQNAFVLSRGIHRNH-HLLAATLCCLCLILIGIVFGGANLAASPIGLAL 71
QY 67 MAWLHPLIMVGGGLYLCWMGYOMLRGALKKE-AVSAPQVELAKSGRSFLKGLLTNLAN 125
D 72 LTV-----GGVFLCFWGRSRSRWSAWOGGAALADSPRLMGVRSVLAITLGV--TLLN 122
QY 126 PKAIIYFGSVFSL-FVGDVNGTGTARWGIFALIIVETLAWFTVVA--SLFALPOMRRGYOR 182
D 123 PH--VYLDITLMLGFGSQRFAEPLRPAFAAGAMLASLVWFYSFAFGAALSPWLARG--R 178
QY 183 LAKWIDGFAGALFAGFGIHL 202
D 179 VQAIDITVIGLIMLGLALQL 198
```

```
RESULT 11
LYSE_CORGL STANDARD; PRT; 236 AA.
AC P94633;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LYSINE EXPORTER PROTEIN.
GN LYSE.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteriidae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI27;
RX MEDLINE=97126810; PubMed=89711704;
RA Vrijic M., Sam H., Eggeling L.;
RT "A new type of transporter with a new type of cellular function: L-lysine export from Corynebacterium glutamicum.";
RL Mol. Microbiol. 22:815-826(1996).
CC -!- FUNCTION: INVOLVED IN THE EFFLUX OF EXCESS OF L-LYSINE. THIS IS NECESSARY TO CONTROL THE INTRACELLULAR L-LYSINE LEVEL.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X96471; CAA65324.1;
DR InterPro; IPR001123;
DR Pfam; PF01810; Lyse; 1.
KW Transport; Transmembrane; Inner membrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
SQ SEQUENCE 236 AA; 25425 MW; E8B6B8CF037C4727 CRC64;

Query Match 10.0%; Score 105; DB 1; Length 236;
Best Local Similarity 20.4%; Pred. No. 0.085;
Matches 51; Conservative 36; Mismatches 99; Indels 64; Gaps 9;

QY 1 MLMLFLTVAMVHVALMSPGDFEFVSQTAVSRKKEAMGVLTGTC---GVNWMAGIALGLHLI-IEK 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 IMEIPITGLLGSALLSIGPQNVLVIQGIKR---EGLINVLVCLISDVFLFIAGTLG 59
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 60 LHL-----IIEKMAWLHTLWVGGLYLCWMGYQMLRGALKKEAVSAPAPV----- 106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 60 VDLLSNAPIVLDIRW-----GGIATLLWFVMAAKDAWTKN---VEAPQIIETEP 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 107 -----ELAKSGRS-----FLKGLLTNLNPK-----AIIYFGS 134
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 110 TVPDDTPLGSSAVATDTRNRVVEVSVDKQVWKPMMLAIVLTWLPNNAYLDAFFVIGG 169
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 VFSLFVDNNGVTTARWIGIFALIVETLAWFTVWASLFAQLPQMRRGYQRLAKWIDGFAGAL 194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 170 -----VGAQVGDTRWIFAAGAFASLLIWFPLVGFGAALSRLSPSPKVRWLNVVAVV 224
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 195 FAGFGIHLII 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 225 MTAIAIKML 234
```

```
RESULT 12
YGGA_AERHY STANDARD; PRT; 225 AA.
AC P52047;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL 24.5 KDA PROTEIN IN AHYR-CDPD INTERGENIC REGION.
OS Aeromonas hydrophila.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OC Aeromonas.
OX NCBI_TaxID=644;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AI;
RX MEDLINE=97431471; PubMed=9286976;
RA Swift S., Kariyhev A.V., Fish L., Durant E.L., Winson M.K., Chhabra S.R., Williams P., Macintyre S., Stewart G.S.A.B.;
RT "Quorum sensing in Aeromonas hydrophila and Aeromonas salmonicida: identification of the luxR homologs AhyR1 and AsaR1 and their cognate N-acylhomoserine lactone signal molecules.";
RL J. Bacteriol. 179:5271-5281(1997).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X89469; CAA61655.1; ALT_INIT.
DR InterPro; IPR001123;
DR Pfam; PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 116 136 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
SQ SEQUENCE 225 AA; 24482 MW; 172DB104473B0B09 CRC64;

Query Match 9.9%; Score 104.5; DB 1; Length 225;
Best Local Similarity 26.1%; Pred. No. 0.089;
Matches 57; Conservative 29; Mismatches 91; Indels 41; Gaps 12;

QY 14 VALMSP-GPDEFFVSQTAVSRKKEAMGVLTGTC---GVNWMAGIALGLHLI-IEK 66
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 13 LAMIIPIGAQNAFVLSRGIRHNH-HLLTATLCCLDVLIGIGVFGGANLLAASPIGLAL 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 67 MAWLHTLWVGGLYLCWMGYQMLRGALKKE-AVSAPAPVELAKSGRFLKGLLTNLN 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 72 LTV-----GGVFLGWFGIRSLRSARWQCGAKLADSPQLMGVKSIVLAMP LGV--TLLN 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 126 PKAIYFSGVPSL-FVGDNGVTTARWIGIFALIVETLAWF-----TVVASLFAIPQMR 178
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 PH--VYLDTLMLLGSFGSQFAEELRSAPNAVAMLASLWVFSYSLAFGAVVLSPW-LARSQ 179
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 GYORL-----AKWIDGFAGALFAGFGIHLI 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 180 GYSKLLILLVSPCWGRCNWRAGLCWRHKAIEVPHLI 217
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
YGGA_ECOLI STANDARD; PRT; 211 AA.
AC F11667;
DT 01-OCT-1989 (Rel. 12, Created)
```

[illegible]

```

RESULT      7
ID           Q9KLA0      PRELIMINARY;      PRT;      204 AA.
OC          ACACLA0;
DT          DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT          DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT          DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE          DE HYPOTHETICAL PROTEIN VCA0846.
GN          VCA0846.
OS          OS Vibrio cholerae.
OC          OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX          OX NCBI_Taxid=666;
[]
RN          RN SEQUENCE FROM N.A.
RC          RC STRAINE=EL TOR N16961 / SEROTYPE O1;
RX          RX MEDLINE=20406833; PubMed=10952301;
RA          RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA          Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA          Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA          Ermolaeva M.D., Vamathevan J., Bess S., Qin H., Dragoi I., Sellers P.
RA          McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
RA          Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA          Fraser C.M.;
RT          RT "pna sequence of both chromosomes of the cholera pathogen Vibrio
RT          cholerae.";
RL          RL Nature 406:477-483(2000).
DR          DR EMBL; AEO04412; AAF96744.1; -.
DR          DR TIGR; VCA0846; -.
SQ          SQ SEQUENCE      204 AA: 21719 MW: D0988729659FEBD CRC64:

```

RESULT 8
Q9RMX0
ID Q9RMX0 PRELIMINARY; PRT; 205 AA.
AC Q9RMX0;
DT 01-MAY-2000 (TRENBLrel. 13, Created)

```

RESULT      5
Q916H7
ID          Q916H7      PRELIMINARY;      PRT;      212 AA.
DT          01-OCT-2000 (TREMBLrel. 15, Created)
DT          01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT          01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE          AMINO ACID EFFLUX-LIKE PROTEIN.
OS          Brucella melitensis.
OC          Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC          Brucellaceae; Brucella.
OX          NCBI_TaxID=29459;
RN          [1]
RP          SEQUENCE FROM N.A.
RC          STRAIN=16M;
RA          Zygmunt M.S., Diaz M.A., Teixeira-Gomes A.P., Cloeckaert A.;
RT          "Cloning, nucleotide sequence, and expression of the Brucella
RT          melitensis suc gene coding for a dihydrolipoamide succinyltransferase
RT          homologous protein.";
RT          Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
SQ          EMBL; AF235020; AAF43702.1; -.
DR          NCBI SEQUENCE 212 AA; 23158 MW; 0C21287CEA665F56 CRC64;

```

```

RESULT      6
Q9KFP9      Q9KFP9      PRELIMINARY;      PRT;      207 AA.
ID          AC
Q9KFP9;    01-OCT-2000 (TremBLrel. 15, Created)
DT         DT
DT         01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT         01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE         BH0429 PROTEIN.
OS         BH0429.
OS         Bacillus halodurans.
OC         Bacteria; Firmicutes; Bacillus/Clostridium group;
OC         Bacillus/Staphylococcus group; Bacillus.
OX         NCBI_TaxId=86665;
RN         [1]
RP         SEQUENCE FROM N.A.
RC         STRAIN=C-125 / JCM 9153;
RA         Takami H., Nakasone K., Takaki Y.;
RL         Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
EMBL;      AP001508; BAB04148.1; -.
DR         EMBL;
SQ         SEQUENCE      207 AA;      22317 MW;      9FED148C8EDD3F0 CRC64;

```

Query Match 19.1%; Score 201; DB 2; Length 207;
Best Local Similarity 28.3%; Pred. No. 2.3e-09;
Matches 63; Conservative 39; Mismatches 81; Indels 40; Gaps 9;
Qy 2 LMLFTNVMVHTIALMSPGDFFTVSOTAVSRSRKEAMGVLGITCCVMVWAGIALGLH 61


```
DR PFAM; PF01810; Lyse; 1.
SQ SEQUENCE 213 AA; 23614 MW; E97242DD507C8065 CRC64;

Query Match 15.5%; Score 163; DB 2; Length 213;
Best Local Similarity 25.3%; Pred. No. 3.3e-06;
Matches 47; Conservative 35; Mismatches 76; Indels 28; Gaps 5;

QY 2 LMLFLTVAMVHVALMSGPDPFFVVSQTAVSRKREAMMGVLGTCGVWVWAGIALLGLH 61
: ||| | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 19 IYFILTVLFSI-----SPGAMMFVQQSQKQKGVKTGLAAVLGTEIGVFYIVILTALGIS 74
QY 62 LIIEKMAWLHTLIMVGGGLYLCWGMQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLT 121
: : : : : | | | | | | | | | | | | | | | | | | | | | | | | |
Db 75 TVLKEPSYITGLQIGRAYLLIYAL-----SWPQNASNQTPTASRSYTGTFIQGLVLI 130
QY 122 NLANPKAIIFYGVSFLFV--GDNVGTARGIF-----ALJ-----IVETL 161
: || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 NLTNPKRIVLFLSLIPQFVPRDSNAMTFVYGLIFNTSGLLVNFVALLADRVNRMLSRV 190
QY 162 AWFTVV 167
Db 191 TWENYV 196

RESULT 14
Q9KJ3 PRELIMINARY; PRT; 208 AA.
AC Q9KJ3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE HYPOTHETICAL PROTEIN VCA0355.
GN VCA0355.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EL TOR N16961 / SEROTYPE O1;
RA MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004372; AAF96263.1; -.
DR TIGR; VCA0355; -.
SQ SEQUENCE 208 AA; 22111 MW; 3CB406979F6D39 CRC64;

Query Match 14.8%; Score 155.5; DB 2; Length 208;
Best Local Similarity 23.8%; Pred. No. 1.3e-05;
Matches 50; Conservative 43; Mismatches 102; Indels 15; Gaps 6;

QY 2 LMLFLTVAMVHVALMSGPDPFFVVSQTAVSRKREAMMGVLGTCGVWVWAGIALLGLH 61
: ||| | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 6 LLLFIVAC---LAINIPGPDVYIYVSNTMKGKLVTFGKAAMGLGVFVHTLAASLGLS 62
QY 62 LIIEKMAWLHTLIMVGGGLYLCWGMQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLT 121
: || | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 63 AILSSAVAFSAVKWLGAAIYVLGVQSLCSMWRGSGTLKSVESVESDKN--VFQGVIV 120
QY 122 NLANPKAIIFYGVSFLFV--GDNVGTAR---W-GIFALIVE--TLAFTVVASLALPQ 175
: ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 SVLNPKRVALFSLFLPQFIDTSSGSASMQLLWGLLFLSVLTMCNILYASGVSWFSRPN 180
QY 176 MRRGYQLAKWIDGFAGALFAGEGHIILIS 205
```

```
Db 181 S-----QRYSRGLEGVSGVLLIGLASKVAIS 206
|| : : : || : | | | : ||
|| : : : || : | | | : ||

RESULT 15
Q9L6N6 PRELIMINARY; PRT; 206 AA.
AC Q9L6N6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE RTHB PROTEIN.
DE RTHB.
OS Salmonella typhimurium LT2.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=99287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SGSCL412;
RA WashU;
RT "The Salmonella typhimurium Genome Sequencing Project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SGSCL412;
RA Waterston R.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
SQ EMBL; AF23324; AAF33432.1; -.
SQ SEQUENCE 206 AA; 22322 MW; E089B357D30750F2 CRC64;

Query Match 14.6%; Score 153.5; DB 2; Length 206;
Best Local Similarity 23.4%; Pred. No. 1.9e-05;
Matches 47; Conservative 38; Mismatches 91; Indels 25; Gaps 5;

QY 17 MSPGPDFFVVSQTAVSRKREAMMGVLGTCGVWVWAGIALLGLHLIEKMAWLHILWV 76
: ||| | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 18 LSPGSGAINTMTTSINHGYRGAVASIAQLQTGLGIHIVLVGVGLTFSRSLAFELKW 77
QY 77 GGGLYLCWGMQMLR--GALKKEAVSAPAPQVELAKSGRSFLKGLLTNLNPKAIIFYGS 134
: ||| | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 78 AGAAVLLIWLGIQOWRAAGIDILHILA-----QTQSRGRLEKRAIFVNLNPKSIYFLAA 131
QY 135 VFSLEFVGDNVGTARWGTFALIIVETLAWFTVVASLFPQMRRGYQRLAKWIDG----- 189
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 132 LFEPQIMPQPOLAQY----LILGVT----TIVYDMVVMVTGYATLAQRIAAWIKGPKQMK 183
QY 190 ----FAGALFAGFGIHLIIS 206
: || | | | | | | | | | | | | | | | | | | | | | | |
Db 184 ALNKAFGSLFMLVGLLASAR 204

Search completed: May 6, 2001, 14:40:51
Job time: 247 sec
```

alignment_block:

US-09-466-935-4 x AW863870/rev ..

Align seg 1/1 to reverse of: AW863870 from: 1 to: 535

5 PheLeuThrValAlaMetValHisIleValAlaLeuMetSerProGlyPr 21
|||||.....
235 TTTTACCGTAGCCTTGATTCACTTGTGGCAGTGGCGAGCCCGCGCC 186

21 oAspPhePheValSerGlnThrAlaValSerArgSerArgLysGlu. 37
|||||.....
185 GGATTTCGCCGTGGTGGCGTGAAGCGTGACCCAGCCGCGCGTGGTG 136

38 AlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaCl 54
|||.....
135 GCACATGGACGGCCCTTGGCGTGGTTCGGCGATTTCCTCATGTGGG 86

54 yIleAlaLeuGlyLeuHisIleIleLeuLys 66
|.....
85 TTTACTGTTGCTGGGTATCGGTTTGATCGTGCCCGAG 49

seq_name: gb_est52:AW863932

seq_documentation_block: 535 bp mRNA EST 22-MAY-2000
LOCUS AW863932
DEFINITION PM4-SN0012-030400-001-g02 SN0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW863932
VERSION AW863932.1 GI:7997982

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 535)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

2020663

Contact: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil.

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=st5-pm4-SN0012-030

400-001-g02&t3=2000-04-03&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 8

High quality sequence stop: 533.

Location/Qualifiers

1..535

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="SN0012"

/dev_stage="Adult"

/note="Organ: stomach_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products

derived from ORSTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

BASE COUNT

134 a 176 c 133 g 92 t

ORIGIN

alignment_scores:
Quality: 118.00 Length: 63
Ratio: 2.458 Gaps: 1
Percent Similarity: 76.190 Percent Identity: 39.683

alignment_block:

US-09-466-935-4 x AW863932/rev ..

Align seg 1/1 to reverse of: AW863932 from: 1 to: 535

5 PheLeuThrValAlaMetValHisIleValAlaLeuMetSerProGlyPr 21
|||||.....
235 TTTTACCGTAGCCTTGATTCACTTGTGGCAGTGGCGAGCCCGCGCC 186

21 oAspPhePheValSerGlnThrAlaValSerArgSerArgLysGlu. 37
|||||.....
185 GGATTTCGCCGTGGTGGCGTGAAGCGTGACCCAGCCGCGCGTGGTG 136

38 AlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaGl 54
|||.....
135 GCACATGGACGGCCCTTGGCGTGGTTCGGCGATTTCCTCATGTGGG 86

54 yIleAlaLeuGlyLeuHisIleIleLeuLys 66
|.....
85 TTTACTGTTGCTGGGTATCGGTTTGATCGTGCCCGAG 49

seq_name: gb_est96:BG116685

seq_documentation_block:

LOCUS BG116685 1305 bp mRNA EST 30-JAN-2001
DEFINITION 602317878F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417986 5',
mRNA sequence.

ACCESSION BG116685

VERSION BG116685.1 GI:12610191

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Life Technologies, Inc.

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLAM10151 row: h column: 19

High quality sequence stop: 182.

Location/Qualifiers

1..1305

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="IMAGE:4417986"

/clone_lib="NIH_MGC_88"

/tissue_type="duodenal adenocarcinoma, cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: small intestine; Vector: pCMV-SPORT6;

Site_1: NotI; Site_2: SalI; Cloned unidirectionally;

oligo-dr primed. Average insert size 1.767 kb. Library

enriched for full-length clones and constructed by Life

Technologies. Note: this is a NIH_MGC Library."

BASE COUNT

457 a 463 c 347 g 38 t

ORIGIN

seq name: qb qss14:AO991044

```

99 valSerAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPh 115
   ||| |||||: : : : : : : : : : : : : : : : : : : : : : : : : : :
312 ...TTTCCCAATATTGAACAGGTGTCAGGCGAGTAGTACAGCAAAAACCG 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
115 eLeuLysGlyLeuLeuThrAsnLeuAla.....AsnPro.... 126
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
265 TTGGCGGGTCATTGCGATTATCTTCGCGTCACCTGGTTAAATCCGCAFG 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
127 .....LysAlaIleIleTyrPheGlySerValPheSerLeuPheVal 140
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
215 TTTATTATTAGACACCATGTTGTTCTGGGCAGT.....ATT 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
141 GlyAspAsnValGlyThrThrAlaArq...TTPGlyIlePheAlaLeuI 156
   ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
180 GCGCGCCAGTTATCCTCAGAACTCAGGCCCTGTTTACTTTGGTCCGC 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
156 eIleValGluThrLeuAlaTrpPhe.....ThrValValAlaSerL 170
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 AAGCGCT...TCGTGATTGTTGTTTTCCTCTGTCATTTGCTAGCGGCAT 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
170 euPheAlaLeuProGlnMetArgGlyTyrGlnArgLeuAlaLysTrp 186
   |||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
83 GGTTCCTCCCTGTTAAGTAAGCGCTTCTCAACGGATT..... 43
   |||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
187 IleAspGlyPhe 190
   |||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
42 ATTAATGGGTTT 31
   |||: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

seq_name: gb_gss27:B07708

seq_documentation_block: 709 bp DNA GSS 15-NOV-1997
LOCUS B07708
DEFINITION 736HIC0930305 Rhodobacter sphaeroides 2.4.1 genomic DNA library
Rhodobacter sphaeroides genomic clone 736HIC0930305 similar to orf
(D64002), DNA sequence.
B07708
ACCESSION B07708.1 GI:2055500
VERSION B07708
KEYWORDS GSS.
SOURCE Rhodobacter sphaeroides.
ORGANISM Rhodobacter sphaeroides
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Rhodobacter.
REFERENCE 1 (bases 1 to 709)
AUTHORS Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.
and Kaplan,S.
TITLE Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
chromosome II is a true chromosome
JOURNAL Microbiology 143, 3085-3099 (1997)
MEDLINE 98015398
COMMENT Contact: Mackenzie, C.
Department of Microbiology and Molecular Genetics
The University of Texas Health Science Center
6431 Fannin St., Houston, TX 77006, USA
Tel: (713) 794-1742
Fax: (713) 794-1782
Email: mackenz@utmsg.med.utu.tmc.edu
Seq primer: pBluescript SK (-) T3
Class: shotgun.
Location/Qualifiers
1..709
/organism="Rhodobacter sphaeroides"
/strain="2.4.1T"
/db_xref="taxon:1063"
/clone="736HIC0930305"
/clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
library"
/lab_host="E. coli SL7-1"
/note="Vector: pLA2917; Genomic DNA from Rhodobacter
sphaeroides was prepared and partially digested with
Sau3AI. Size selected (20kb) fragments were subcloned into
the BglII site of the cosmid vector pLA2917 (Allen, L. N.
and R. S. Hanson, 1985. Journal of Bacteriology,
161:955-962. The library was then ordered around

```

```

chromosome II of Rhodobacter sphaeroides (Choudhary et
al., 1994. Journal of Bacteriology, 176:7694-7702). The
cosmids were then digested with a variety of restriction
enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and
DNaseI) and restriction fragments subcloned into the
respective multiple cloning site sites of pBluescript SK
(-). Note BglII fragments were subcloned into the
pBluescript BamHI site. DNaseI fragments were subcloned
into the EcoRV site. All subclones were transformed into
E. coli XLiBlue MRF+. All fragments were then sequenced
and the sequences where possible were assembled using the
CGC program GELASSEMBLE.
BASE COUNT 125 a 200 c 210 g 166 t 8 others
ORIGIN

alignment_scores:
Quality: 90.50 Length: 79
Ratio: 1.926 Gaps: 1
Percent Similarity: 59.494 Percent identity: 29.114

alignment_block:
US-09-466-935-4 x B07708 ..
Align seg 1/1 to: B07708 from: 1 to: 709

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProgl 20
|||||: : : : : ||| ||||| : : : : : |||||
467 CTCTTCATGTTGCTGCTCGGGTG...ATCGTTCTGACNATCAGCCCGG 513

20 YProAspPhePheValSerGlnThrAlaValSerArgSerArgLysg 37
|||||: : : : : : : : : : : : : : : : : : : : : |||||
514 CCGCGATCTTTTCGTCATCTCGGCGCGGCATCTCAAGTCGGCGCG 563

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
|||||: : : : : : : : : : : : : : : : : : : : : : :
564 CCGCCTGTTCACCGCTTTCGGGCTTCTTCGGCGGCTTCATCAAGTG 613

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLe 70
: : : : : ||||| : : : : : : : : : : : : : : : |||
614 CCCTCTGCTTGGCTTGGAGTTCGCNACGCTNTTGTGCGCAGCCCGCT 663

70 uHisThrLeuIleMetValGlyGlyGlyLeuTyrLeu 82
| : : : : ||||| |||||
664 TTTCGACATCTCGCTGCTTATGGGGTGGTGCATATCTT 700

seq_name: gb_est79:BB872275

seq_documentation_block: 1060 bp mRNA EST 20-OCT-2000
LOCUS BB872275
DEFINITION 601446314F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850491 5',
mRNA sequence.
ACCESSION BB872275
VERSION BB872275.1 GI:10321051
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1060)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
DNA distribution: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```



```
Quality: 90.50 Length: 219
Ratio: 0.943 Gaps: 10
Percent Similarity: 43.836 Percent Identity: 23.744

alignment_block:
US-09-466-935-4 x BE795531 ..

Align seg 1/1 to: BE795531 from: 1 to: 1183

16 LeuMetSerProGlyProAspPhePheValSerGlnThrAlaValSe 32
||||| ||| :|| :||| :||| :||| :||| :||| :|||
411 CTGAGCTGAGGGGAGTGAATTTGAACCTTCTGACCGAGTCGAGTGAG 460

32 rArgSerArgLysGluAlaMetMetGlyVal.....L 43
|:|:|:| | | | | | | | | | | | | | | | | | |
461 TCAGGCTGGACCTAGCAGTGTGAGGCCACTGAGCGAGAGATCTCTT 510

43 euGlyIleThrCysGlyValMetValTrpAlaGlyIleAlaLeuGly 59
:| | | | | | | | | | | | | | | | | | | | | |
511 TTGGGAGTCACCTGTGGT.....TTAGGT 533

60 LeuHisLeuIleIleGluLysMetAlaTrpLeuHisThrLeuIleMetVa 76
:| | | | | | | | | | | | | | | | | | | | | |
534 TTTTCTTCCCAATC.....CCACTTAAGGTGAT 562

76 lGlyGlyGly.....LeuTyLeuCysT 84
:| | | | | | | | | | | | | | | | | | | | | |
563 GGGGCAAGGAAGAACTACAGCTCCCTCCCTCAATATATTTGTATGCT 612

84 rPMetGlyTyGlnMetLeuArgGlyAlaLeuLysGluAlaValSer 100
| | | | | | | | | | | | | | | | | | | | | |
613 GGGTGGCTGGCTGAGCTCAGAGGCTCTTAGAGAGGACACTACCTCC 562

101 AlaProAlaProGlnVal.....G 107
| | | | | | | | | | | | | | | | | | | | | |
663 CCTCCCAACAGCTGGATGCCATTTCTGAGCTCAAGTCACTGAAGTGAGA 712

107 uLeuAlaLysSerGlyArg..... 113
| | | | | | | | | | | | | | | | | | | | | |
713 GTGTGCTCCCAAGGAGGCGCTCTCTCCATCAGGATGGGTACTGTGGG 762

114 .....SerPheLeuLysGlyLeuLeuThr 121
| | | | | | | | | | | | | | | | | | | | | |
763 GGAACAAATAGTCAGGCTATTGGTTCCCTTTTGAGAGTGTCTGCGAA 812

122 AsnLeuAlaAsn.....ProLysAlaIleIleTy 131
| | | | | | | | | | | | | | | | | | | | | |
813 GCACTAAAGATGATGTCAGGCGGCTCCTCTGAAAGACATGCTG... 860

131 rPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThr.... 146
:| | | | | | | | | | | | | | | | | | | | | |
861 ....GCTTCTACTATTTCGCTCTCTAGAGGCCGCCAAGGTCCGACGAGCT 906

147 ..ThrAlaArgTTPGlyIlePheAlaLeuIleIleValGluThrLeuAla 162
| | | | | | | | | | | | | | | | | | | | | |
907 GGACAGTACTGCGCCACCTCTCTCAGAGCAAGTCCGCTTGTGGCT 956

163 TrpPheThrValValAla.....SerLeuPheAlaLeuProGlnMetAr 177
:| | | | | | | | | | | | | | | | | | | | | |
957 AGATGGCGGTAGCGAGGATAGGATCTACAAAGGTTGCCACGCGCGGA 1006

177 qArgGly 179
:| | | | | | | | | | | | | | | | | | | | | |
1007 AAGGGGT 1013
```

seq_name: gb_est94:BF972121

seq_documentation_block:

LOCUS BF972121 1749 bp mRNA

DEFINITION 602240161F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4328918 5',

ACCESSION BF972121

VERSION BF972121.1 GI:12339336

KEYWORDS
SOURCE
ORGANISMEST.
human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 1749)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LLCMI189 row: i column: 15

High quality sequence stop: 329.

Location/Qualifiers

1..1749

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:4328918"

/clone_lib="NIH_MGC_46"

/tissue_type="leiomyosarcoma cell line"

/lab_host="DH10B (phage-resistant)"

/note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2:

EcoRI; cDNA made by oligo-dT priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGCAG(6). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH_MGC

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

Library."

BASE COUNT 623 a 603 c 373 g 150 t
ORIGIN

alignment_scores:

Quality: 90.50 Length: 172

Ratio: 1.104 Gaps: 12

Percent Similarity: 47.674 Percent Identity: 24.419

alignment_block:

US-09-466-935-4 x BF972121/rev ..

Align seg 1/1 to reverse of: BF972121 from: 1 to: 1749

41 GlyValLeuGly...IleThrCysGlyValMetValTrpAlaGlyIleAl 56

||||| ||| :||| :||| :||| :||| :||| :||| :|||

855 GGTGTGTGGCGCGGTGGCGGCGCTTTGTTCTATGACGCTCTTC 806

56 aLeuLeuGlyLeuHis.....LeuIleIleGluLysMetAlaTrpLeuH 71

:| | | | | | | | | | | | | | | | | | | | | |

805 GGTCGTGGGTGTGTTCTCTAGTTGGTGT.....TGTGGGTCT 765

71 isThrLeuIleMetValGlyGlyLeuTyLeuCysTrpMetGly... 86

:| | | | | | | | | | | | | | | | | | | | | |

764 TTTCTGTGTAGTCGTGGGGGACCCCTATGG...TGTGGAGTGGGCG 718

87TyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSe 100

||| :||| :||| :||| :||| :||| :||| :||| :|||

717 TTTGTGTGTACTATTATGTCGTGGGTCTCTTCGCGG...GCGGTAC 671

100 rAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuL 117

:| | | | | | | | | | | | | | | | | | | | | |

670 AGCGTCGCGCCCTGTGCTC..... 652

117 ysglyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyPheGly 133

165 hrValValAlaSerLeuPheAlaLeuProGlnMet 176
 :::::::::::::::::::: ::::::::::::::
 243 GCATCACCTGCAGTCTGCTCAGGTTCCCGAGGCTC 209